



DRAWING MORALS

Essays in Ethical Theory

THOMAS HURKA

Drawing Morals

Oxford Moral Theory

Series Editor

David Copp, University of Florida

Drawing Morals: Essays in Ethical Theory

Thomas Hurka

Against Absolute Goodness

Richard Kraut

Drawing Morals

ESSAYS IN ETHICAL THEORY

Thomas Hurka

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2011 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
www.oup.com

Oxford is a registered trademark of Oxford University Press.

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Hurka, Thomas, 1952–

Drawing morals : essays in ethical theory / by Thomas Hurka.
p. cm.

Includes bibliographical references and index.

ISBN 978-0-19-974309-4 (acid-free paper) 1. Ethics. I. Title.
BJ21.H87 2011
170—dc22

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

{ CONTENTS }

Acknowledgments	vii
Introduction	ix

PART I Methodology

1. Normative Ethics: Back to the Future	3
---	---

PART II Comparing and Combining Goods

2. Value and Population Size	25
3. The Well-Rounded Life	37
4. Monism, Pluralism, and Rational Regret	55
5. How Great a Good Is Virtue?	74
6. Two Kinds of Organic Unity	94
7. Asymmetries in Value	113

PART III Individual Goods

8. Why Value Autonomy?	139
9. Desert: Individualistic and Holistic	154
10. Virtuous Act, Virtuous Disposition	178
11. Games and the Good	185

PART IV Principles of Right

12. Rights and Capital Punishment	201
13. Two Kinds of Satisficing	215
14. The Justification of National Partiality	219
15. Proportionality in the Morality of War	238
References	265
Index	272

This page intentionally left blank

{ ACKNOWLEDGMENTS }

I am grateful to Peter Ohlin of Oxford University Press for suggesting this volume, to David Copp for welcoming it into his series, and to both for helpful advice on what to include. Thanks also to Terry Teskey for copyediting both now and over the years, and to all the many philosophers with whom I've discussed the ideas in these essays or whose work has influenced mine.

These essays originally appeared in the places listed below. The author and publisher gratefully acknowledge permission to reprint them here.

"Normative Ethics: Back to the Future," in Brian Leiter, ed., *The Future for Philosophy* (Oxford: Clarendon Press, 2004), pp. 246–64.

"Value and Population Size," *Ethics* 93 (1983): 496–507.

"The Well-Rounded Life," *The Journal of Philosophy* 84 (1987): 727–46.

"Monism, Pluralism, and Rational Regret," *Ethics* 106 (1996): 555–75.

"How Great a Good Is Virtue?" *The Journal of Philosophy* 95 (1998): 283–304.

"Two Kinds of Organic Unity," *The Journal of Ethics* 2 (1998): 283–304.

"Asymmetries in Value," *Nous* 44 (2010): 199–223.

"Why Value Autonomy?" *Social Theory and Practice* 13 (1987): 361–82.

"Desert: Individualistic and Holistic," in Serena Olsaretti, ed., *Desert and Justice* (Oxford: Clarendon Press, 2003), pp. 45–68.

"Virtuous Acts, Virtuous Dispositions," *Analysis* 66 (2006): 69–76.

"Games and the Good," *Proceedings of the Aristotelian Society, Supplementary Volume* 80 (2006): 217–35.

"Rights and Capital Punishment," *Dialogue* 21 (1982): 647–60.

"Two Kinds of Satisficing," *Philosophical Studies* 59 (1990): 107–11.

"The Justification of National Partiality," in Robert McKim and Jeff McMahan, eds., *The Morality of Nationalism* (New York: Oxford University Press, 1997), pp. 139–57.

"Proportionality in the Morality of War," *Philosophy and Public Affairs* 33 (2005): 34–66.

This page intentionally left blank

{ INTRODUCTION }

This book contains a selection of my essays in moral and political philosophy published between 1982 and 2010. They address a variety of topics. Many concern which states of affairs are intrinsically good, but others discuss which acts or policies are right. Some have abstract topics, such as the principle of organic unities or the nature and value of virtue; others tackle applied issues such as criminal punishment, nationalism, and the use of force in war.

But the essays are also thematically unified. They all address normative topics; none is primarily about metaethics or some particular historical philosopher. And they share a common methodology. They all explore the internal structure of some moral view, asking, for example, what makes some achievements more valuable than others or how different goods or right-making characteristics weigh against each other. More specifically, they all practice what the first essay, “Normative Ethics: Back to the Future,” calls “structural” as against “foundational” moral theory. Structural theory does not try to ground views about what is good or right in claims about some different and supposedly more fundamental topic, such as the agent’s flourishing or the presuppositions of rational agency. Instead, it takes some moral claims to be underderivatively true and proceeds to analyze their content, which it often finds more complex and interesting than foundational theory does. It is the details of a moral view rather than some grand external justification of it that are its primary focus.

The book opens with “Normative Ethics” because that essay gives a general description and defense of the structural method. This allows the essays that follow to be read as illustrating that method and demonstrating its merits; it also clarifies those essays’ ambitions. It is not that I first became convinced that this is the best approach to normative theory and then started writing papers that use it; nor did I first write the papers and only later work out a methodology to fit them. The two developed simultaneously. I was first attracted to ethics in an undergraduate seminar that included work by G.E. Moore and W.D. Ross, whom I think of as exemplifying the structural approach; I was also influenced by graduate seminars of Derek Parfit’s on what would become *Reasons and Persons*. And I was always skeptical of the more pretentious claims of Kantian ethics and the high-mindedness of much writing inspired by Aristotle. So when I began writing my own essays in normative theory they tended to follow the structural line, though I was not then conscious of their doing so. As I continued to write, I became more aware of my methodological views, both positive and negative, and finally expressed them in the piece that is the first chapter here.

One tool of the structural method, used in several of the essays here, is graphs, which try to express moral ideas in visual form. This tool adds little to the analysis of a simple moral view, such as one that equates the overall good in a population with the sum or average of the goods in its members. But other views are more complex. They can say the value of increases in a given good is not constant but diminishes the more of that good there already is, eventually approaching zero. Or they can say a good like virtue or desert is governed by an ideal of proportionality, whereby the best division of virtuous concern between goods is proportioned to their degrees of value, or the best division of rewards or punishments among people matches their degrees of merit or demerit. Here a graph gives an especially clear representation of the view in question, and in particular of the mathematical relations between its parts. (Both diminishing value and proportionality generate curves with a distinctive shape.) The representation does not replace one in words; it only supplements it. But the supplement is in a different and illuminating form.

Graphs have a further merit, since to complete one you often have to address philosophical issues you might otherwise not have thought of. You may be drawing curves with a certain shape, to express a view about, say, virtue or desert. But these curves cut the y-axis and you have to decide where they do so: at the origin, above it, or below it? Or they may have separate parts, for example because they rise to a peak representing a maximum value and then turn down. Here you have to decide whether the curves slope more steeply before or after their peak, and you also have to decide whether the peaks are higher in one part of the graph than another. These questions all raise ethical issues—whether certain outcomes are neutral in value, good, or evil; whether one of two moral failings is more serious; and whether different optimal outcomes have different values—and you need to resolve these issues if you are to understand the view completely. They can all be discussed verbally, apart from any graph. But whereas you might well miss them if you thought only in words, the visual representation forces them upon you. Though just a technical device, it stimulates philosophical reflection.

Any curve on a graph will express a mathematical formula, but it is not always important to know what that formula is: any one yielding a curve with the right general shape will do. Sometimes, however, this is important. There may be a set of intuitively attractive conditions that you want a given moral view to satisfy, and it may not be obvious whether they are all consistent with each other. For example, you may want a view about desert to yield a set of curves that satisfy a proportionality condition, reach a peak on one or other side of the origin, and cut the y-axis at progressively lower points below the origin. To know whether all this is possible, you need to state the conditions mathematically and see if there is a formula that satisfies them all. (There is.) And doing so can have further benefits. It turns out to follow from the proportionality condition that giving a very vicious person his ideally deserved punishment has more value than giving a slightly vicious person his ideal punishment—the peak on the former's curve is higher up. This is intuitively attractive: surely there is a stronger demand of justice to give a serial murderer his deserved

punishment than to do the same to a petty thief. And the fact that it follows from a not obviously connected but equally intuitive condition about proportionality shows that our everyday thinking about desert has an impressive internal unity.

Though a graph can display the internal structure of a view about, say, desert, and highlight the relationships between its parts, it cannot give a deeper justification for the view's having those parts in the first place. It cannot explain *why* desert involves an ideal of proportionality by deriving that ideal from claims about the supposedly more basic topics foundational moral theory appeals to, and it is hard to see how such a derivation could succeed. How could claims about flourishing or rational agency mandate a view about desert with that specific mathematical property? Even if they could, it would surely still be important to describe the view as clearly as possible, so we know exactly what features of it need justifying. That is what a desert graph does, and what the various essays in this book try to do for the different moral ideas they discuss. There are riches in everyday moral thought, and more complexity than philosophers often credit; the essays to follow try to reveal some of those riches.

After the opening methodological chapter, the essays are divided into three sections, two on the good and one on the right. Within each section the essays appear in chronological order, with those published earliest coming first.

This page intentionally left blank

PART I

Methodology

This page intentionally left blank

Normative Ethics

BACK TO THE FUTURE

The course of normative ethics in the twentieth century was a rollercoaster ride, from a period of skilled and confident theorizing in the first third of the century, through a virtual disappearance in the face of various forms of skepticism in the middle third, to a partial revival, though shadowed by remnants of that skepticism, in the final third. The ideal future of normative ethics therefore lies in its past. It must entirely shed its traces of mid-century skepticism if it is to return to the levels of insight provided by G. E. Moore, Hastings Rashdall, J. M. E. McTaggart, W. D. Ross, C. D. Broad, and other early twentieth-century moral theorists.

These theorists shared several fundamental assumptions about ethics, many derived from late nineteenth-century philosophers such as Henry Sidgwick. They were all moral realists, believing that moral judgments are objectively true or false. More importantly, they were nonnaturalist realists, believing, as antirealists also can, that moral judgments form a separate category of judgments, neither reducible to nor derivable from other judgments. For them the property of goodness is not identical to any physical or natural property, and no “ought” can be derived from an “is.” They therefore accepted a realist version of the autonomy of ethics: at the level of fundamental principles, moral judgments are independent of all other judgments.

These theorists also shared the general normative project of systematizing commonsense morality, or finding more abstract principles that can unify and explain our particular judgments about good and right. This project was not unique to them, but their approach to it was distinctively shaped by their belief in the autonomy of ethics.

First, they trusted intuitive judgments as at least a reliable starting point for moral inquiry. Precisely because they denied that moral claims can be derived from other claims, they thought direct intuitive insights, either their own or those of common sense, were the best available entree to the moral realm. They did not see these insights as infallible; they recognized that our intuitions can be

distorted by self-interest and other factors. Nor did they think the disorganized and even conflicting collection of judgments that make up commonsense morality was in that form acceptable. This was their prime motive for theorizing common sense: only if its judgments could be systematized by a few fundamental principles would they be properly scientific. But one test of these principles was their consistency with everyday moral beliefs, and another was their own intuitive appeal. Many of these theorists considered intuitions about general principles more reliable than ones about particular cases, but at all levels of generality they thought the only route to moral knowledge was by some kind of intuitive insight.

Second, and because of their confidence in intuitive judgments, these theorists were in principle open to the whole range of moral views accepted in Western culture.¹ In practice their openness often had a limitation. Many of them were consequentialists, believing that what is right is always what will produce the most good, and as a result they did not say much about nonconsequentialist views. Even Ross, who defended nonconsequentialism, described more its general structure than its details. But Moore, Rashdall, and the others did collectively address a huge variety of views about the good: not only that pleasure is good, but also views valuing knowledge, aesthetic contemplation, virtue, love, and more. Theirs was a golden age of value theory, in part because its theorists defended so many views about what is intrinsically worth pursuing. In addition, they were prepared to explore the details of these views. They did not rest with general claims about their preferred values but produced subtle analyses of the elements of aesthetic contemplation (Moore), the specific character of love (McTaggart), and the forms of virtue and their comparative values (Ross).²

Finally, the analyses these theorists produced were what I will call structural rather than foundational. They described the underlying structure of commonsense judgments and in that sense unified and explained them. But they did not try to ground those judgments in others that concern a more fundamental topic and can therefore justify them to someone who does not initially accept them. Their analyses stayed within a circle of commonsense concepts rather than connecting them to others, either moral or nonmoral, that they saw as more secure. For example, Sidgwick grounded utilitarianism in the principles that one should not prefer a lesser good at one time to a greater good at another or a lesser good for one person to a greater good for another. These principles make explicit two forms of impartiality central to utilitarian thinking; they also help to unify utilitarian claims. But they use similar concepts to those claims rather than relating them to others that are more funda-

¹ They should also have been open to ideas accepted in other cultures, but tended to believe, like many in their time, in the higher moral development of the West.

² G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), pp. 189–202; J. M. E. McTaggart, *The Nature of Existence*, vol. 2 (Cambridge: Cambridge University Press, 1927), pp. 147–61, 436–9; W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 155–73.

mental.³ The same holds for Moore's formulation of the retributive theory of punishment in terms of his principle of organic unities, which says the value of a whole need not equal the sum of the values of its parts. In the case of punishment, Moore argues, a person's having a vicious character is bad, as is his suffering pain, but the combination of a vicious character and pain in the same life is good as a combination, and sufficiently good that inflicting the pain makes the overall situation better.⁴ This analysis again uncovers the structure of retributive claims and unifies them with others that involve organic unities. It also has important implications, for example, that while deserved punishment is good as deserved it is bad as involving pain, so the morally best response to it mixes satisfaction that justice is being done with regret at the infliction of pain. But the analysis does not ground retributivism in some other, less contentious claim and will not persuade someone initially hostile to retributivism. Or consider Broad's treatment of what he calls self-referential altruism. This view holds, contrary to Sidgwick, that our duty to others is not to treat them impartially but to care more about those who are in various ways closer to us, such as our family, friends, and compatriots. Broad's analysis unifies a variety of commonsense claims about the demands of loyalty and invites further inquiry about exactly which relations make for closeness of the relevant kind. But it does not justify self-referentiality in other terms; instead, it assumes self-referentiality, saying that each person should care more about his family and friends because he should in general care more about those who stand in special relations *to him*.⁵

These theorists were aware of more ambitious normative projects: many of their contemporaries proposed deriving moral claims from associationist psychology, Darwinian biology, or Idealist metaphysics. But Sidgwick, Moore, and the others gave both general arguments against this kind of derivation and specific critiques of their contemporaries' views. For them the foundational approach to ethics was illusory and structural analysis the only profitable route to pursue. They did not address every important topic in ethics or leave nothing to be said about those they did discuss. But methodologically they provided a model for how normative inquiry should proceed. And then, in the middle decades of the century, normative ethics virtually disappeared from philosophy.

One cause of this disappearance was the replacement of the moral realism that had dominated the earlier period by crude versions of expressivist antirealism, which held that moral judgments are not true or false but only express pro or con-attitudes, and which understood normative argument as an attempt to transmit these attitudes to others by a kind of emotional contagion.⁶ Another cause was a general conception of philosophy as a second-order discipline, which

³ Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907), pp. 379–84.

⁴ Moore, *Principia Ethica*, pp. 214–16.

⁵ C. D. Broad, "Self and Others," in *Broad's Critical Essays in Moral Philosophy* (London: George Allen & Unwin, 1971).

⁶ See A. J. Ayer, *Language, Truth, and Logic* (London: Gollancz, 1936), ch. 6; and Charles L. Stevenson, *Ethics and Language* (New Haven: Yale University Press, 1944).

analyzes the logic or language of first-order disciplines but does not participate in them itself. Just as philosophy of science analyzes the logic of scientific confirmation but does not itself make scientific claims, so philosophical ethics should study the language of morals but leave substantive moralizing to preachers and poets.

Over time, however, the influence of these causes faded and the intrinsic interest of normative questions, both theoretical and particular, was able to reassert itself. The result was a revival of normative philosophizing in the last third of the century, stimulated especially by the 1971 publication of John Rawls's *A Theory of Justice*.⁷ Since then normative ethics has been a prominent part of the discipline and has produced much valuable work. But its revival has been only partial, held back by remnants of the mid-century's two dominant general attitudes to philosophy.

The first of these was the technical, scientizing attitude of logical positivism and successor views such as W. V. O. Quine's naturalism. This attitude was hostile to common sense and philosophies that take it seriously, holding that the everyday view of the world is riddled with errors. In the linguistic terms that were popular in this period, it held that ordinary language is inadequate for understanding reality and needs to be replaced by a more scientific language, such as first-order logic. The second attitude, which arose in reaction to the first, informed the ordinary-language philosophy of Ludwig Wittgenstein, J. L. Austin, and others. It had a high regard for common sense, which it took to be the repository of centuries of human learning. But it was resolutely antitechnical and antitheoretical. It is a mistake, its partisans held, to think that whenever a single word applies to a set of objects there must be some single property they share; there may be only a loose set of "family resemblances." The attempt to capture that property in an abstract principle does not deepen the insights of common sense but gets them entirely wrong.

Even apart from any general skepticism about philosophical ethics, this pair of attitudes left little room for the earlier project of systematizing commonsense morality. On one side was a view friendly to abstract principles but hostile to common sense; on the other was a view friendly to common sense but hostile to abstract principles. And these two views have continued to influence normative ethics since its re-emergence.

For its part, the scientizing attitude has encouraged philosophers to reject many commonsense moral views as confused or in some other way unacceptable. The result, especially early in the normative revival, was that theoretical ethics considered only a small number of options: utilitarianism, Kantian deontology, and little else. Ideas about desert, natural rights, and virtue were commonly if not universally set aside or analyzed in other, allegedly less suspect terms; either way, the distinctive approaches to ethics they express were ignored. Another object of skepticism was the topic most discussed at the start of the century: intrinsic value. Many philosophers found the idea that there are goods a person should pursue for herself

⁷ John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971).

other than pleasure or the satisfaction of her desires deeply problematic, and they were equally hostile to proposed goods that span people's lives, such as distributions proportioned to their merits, or that lie outside them, such as complex ecosystems. This meant that consequentialism, which in the earlier part of the century had encompassed a wide variety of positions, was in the later period mostly equated with simpleminded versions of utilitarianism. This again cut philosophical ethics off from everyday moral thought. Just as many nonphilosophers use desert and rights as basic moral concepts, so many think there are goods worth pursuing other than pleasure or satisfaction. If a person is compassionate or generous, that in itself makes her life better; likewise if she has deep personal relationships or accomplishes difficult goals. (Consider the longtime recruiting slogan of the U.S. armed forces: "Be all that you can be." It clearly appealed to widespread views about the intrinsic value of developing one's talents.) In rejecting all such views philosophers were again rejecting much of their culture's moral life.

In some cases the grounds of this rejection were conceptual, with philosophers urging, for example, that claims about intrinsic value are simply unintelligible. These particular arguments had little merit. If morality can tell people to pursue others' happiness regardless of whether this will satisfy their own desires, as most philosophers allow, why can it not also tell them to pursue knowledge? And if the worry was that claims about intrinsic value presuppose a suspect realism, that too was groundless. Even if moral realism is problematic, as is by no means clear,⁸ claims about value can always be understood in an expressivist way. In fact, sophisticated versions of expressivism allow virtually any moral view to be accommodated in a scientizing or naturalistic picture of the world. These versions take the attitudes expressed by moral judgments to have the logical form of categorical imperatives, so they are directed at acts or states of a person in a way that is not conditional on that person's having any attitudes. Consider the judgment that malice is evil. According to sophisticated expressivism, someone who makes this judgment expresses a negative attitude to all malice, whatever the malicious person's attitude to malice. If she contemplates malice in someone who has no negative attitude to malice, her own attitude to the malice is negative; if she contemplates a possible world in which she herself has no negative attitude to malice and is malicious, her attitude (from this world) to her malice is negative.⁹ As so formulated, expressivism makes virtually no difference to the study of normative questions. Both realists and expressivists can accept almost any substantive moral view and argue for it in the same way: by appealing to intuitive judgments, formulating abstract principles, and so on. The realists will interpret the judgments as providing insights into moral truth, the expressivists as expressing attitudes they hope others do

⁸ For a recent defense of realism see Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), ch. 8.

⁹ This point is made in recent defenses of expressivism such as Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Mass.: Harvard University Press, 1990); and Simon Blackburn, *Ruling Passions: A Theory of Practical Reasoning* (Oxford: Clarendon Press, 1998).

or can come to share. But viewed on their own, their moral positions will be indistinguishable.

In other cases the grounds for rejecting commonsense views were normative: that people must be free to determine the content of their own good, as desire theories allow, or that all goods are, substantively, goods of individuals. But these arguments were often accompanied by a general sense that the views they targeted were too extravagant to be taken seriously. It is part of the self-image of scientizing philosophers to be tough-minded debunkers of confused folk beliefs, and in ethics this meant rejecting all views with more than a very austere content.¹⁰ But it is again hard to see a persuasive rationale for this approach. Whether moral judgments report a distinctive kind of truth or merely express attitudes, why should their content be limited in this way? Whatever the merits of conceptual parsimony in other domains, it is hard to find one here.¹¹

Relatedly, and especially when ethics was first re-emerging, philosophers tended to confine their attention to structurally simple views rather than recognizing the complexities their predecessors had noted. Whereas Broad had analyzed the structure of self-referential altruism, several prominent works took the main views needing discussion to hold that people should either care only about their own good or care impartially about the good of all.¹² And a common form of argument assumed that if a moral factor such as the difference between killing and allowing to die makes no difference in one context, it cannot make a difference in any context.¹³ But the point of Moore's doctrine of organic unities had been precisely that the difference a factor makes can vary from context to context, depending on what other factors it is combined with.

These scientizing influences have been countered but also complemented by those of the antitheoretical, Wittgensteinian attitude. Its adherents are not debunkers; they are open to many commonsense moral views, especially about the virtues and vices. But they deny that these views can be systematized or captured in abstract principles. Many cite Aristotle's remark that in ethics one should not seek more precision than the subject matter allows,¹⁴ which they take to imply that theorizing

¹⁰ Compare Bernard Williams's claim that utilitarianism's popularity rests on its being a "minimum commitment morality"; see his *Morality: An Introduction to Ethics* (New York: Harper & Row, 1972), p. 91.

¹¹ Note that austerity is not just a matter of theoretical simplicity. Robert Nozick's view that all values are instances of "organic unity" (see his *Philosophical Explanations* [Cambridge, Mass.: Harvard University Press, 1981], pp. 415–28) is theoretically simple but would be rejected by scientizers as hopelessly extravagant.

¹² See, e.g., Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970).

¹³ For the most famous illustration of this argument see James Rachels, "Active and Passive Euthanasia," *New England Journal of Medicine* 292 (Jan. 9, 1975): 78–80.

¹⁴ Aristotle, *Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984), 1094b12–28. Note that, despite this remark, Aristotle gives a mathematical formula for distributive justice at *Nicomachean Ethics* 1131a30–b16, saying that justice requires the ratio of person A's award to B's to equal the ratio of A's merits to B's. He is nothing like the consistent antitheorist some of his contemporary admirers take him to be.

about ethics is fundamentally misguided. The most extreme formulation of their view holds that moral knowledge always concerns particular acts in all their specificity; it does not generalize to other acts, cannot be codified in general principles, and is a matter only of trained moral insight.¹⁵ People with the right moral character can “see” what is right, just, or virtuous in a given situation, but they cannot express that vision in other terms or communicate it to those who lack it.

In my view an antitheoretical position is properly open only to those who have made a serious effort to theorize a given domain and found that it cannot succeed. Antitheorists who do not make this effort are simply being lazy, like Wittgenstein himself. His central example of a concept that cannot be given a unifying analysis was that of a game,¹⁶ but in one of the great underappreciated books of the twentieth century Bernard Suits gives perfectly persuasive necessary and sufficient conditions for something’s being a game. (Roughly: in playing a game one pursues a goal that can be described independently of the game, such as directing a ball into a hole in the ground, while willingly accepting rules that forbid the most efficient means to that goal, such as placing the ball in the hole by hand.) With an exemplary lightness of touch, Suits mentions Wittgenstein only once:

“Don’t say,” Wittgenstein admonishes us, “there must be something common or they would not be called ‘games’”—but *look and see* whether there is anything common to all.” This is unexceptionable advice. Unfortunately, Wittgenstein himself did not follow it. He looked, to be sure, but because he had decided beforehand that games are indefinable, his look was fleeting, and he saw very little.¹⁷

Similarly, ethical antitheorists have decided beforehand that there can be no unifying account of, say, the human good, and therefore do not try seriously to construct one. More specifically, they typically consider only simpleminded ethical analyses and take the failure of those to demonstrate the impossibility of all analyses. But, for example, the fact that not all pleasure is good, because sadistic pleasure is bad, does not refute all general theories of the value of pleasure. An only slightly complex theory can say that sadistic pleasure, while good as pleasure, is bad as sadistic, and more bad than it is good; a more complex theory can say that when pleasure is sadistic it loses its goodness as pleasure. Far from abandoning generality, these analyses use it to illuminate values in a way antitheorists never could.

¹⁵ See, e.g., John McDowell, “Virtue and Reason,” *The Monist* 62 (1979): 331–50; and Jonathan Dancy, *Moral Reasons* (Oxford: Blackwell, 1993), chs. 4–6.

¹⁶ Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1972), sec. 66.

¹⁷ Bernard Suits, *The Grasshopper: Games, Life and Utopia* (Toronto: University of Toronto Press, 1978), p. x. Note that Suits’s analysis of games is structural; once it is in hand, Wittgenstein’s fussing about the differences between games that use boards and cards or that are and are not amusing is embarrassingly superficial.

Despite their differences, scientizers and antitheorists share a common assumption: that any acceptable moral theory must be simple in its content and form. Believing in theory, the scientizers confine their attention to simple views; finding those views unacceptable, antitheorists abandon theory. But both are hostile to the systematic analysis of complex moral views that had been the hallmark of the earlier period. Consider, for example, the topic of how different moral considerations weigh against each other. Scientizers are suspicious of such weighing, especially if it rests on intuitive judgments. They want a mechanical and even empirically implementable procedure for weighing values, as there would be if all values reduced to a single one.¹⁸ By contrast, antitheorists embrace a plurality of values but insist they are “incommensurable,” which they take to imply that nothing systematic can be said about how they compare.¹⁹ The early twentieth-century theorists avoided both these extremes. While recognizing that different values cannot be weighed precisely, they insisted that we can make rough comparative judgments about them, such as that an instance of one good is much, moderately, or only a little better than an instance of another. They also pursued structural questions, such as whether the complete absence of one good—say, knowledge—can always be compensated for by a sufficient quantity of another. Theirs was the intermediate approach of partly theorizing the partly theorizable, but it is excluded by both the scientizing and antitheoretical attitudes.

Among some ethicists the influence of these attitudes is now fading.²⁰ In the last decade or so there has been more sympathy for views that would earlier have been rejected as extravagant, such as perfectionist accounts of each person’s good,²¹ as well as detailed explorations of nonpersonal goods such as equality²² and desert.²³ There has also been greater awareness of the complexity moral views can have; thus, the point that a factor’s importance can depend on its relations to other factors has been made and is widely accepted.²⁴ Scientizing and antitheoretical

¹⁸ Rawls expresses a moderate version of this view in *A Theory of Justice*, saying that pluralist views that weigh principles intuitively are less satisfactory than ones that give some principles lexical priority over others; so the former’s demands always take precedence over the latter’s (pp. 40–45).

¹⁹ See, e.g., Martha C. Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (Cambridge: Cambridge University Press, 1986), pp. 107–17, 294–98; and Michael Stocker, *Plural and Conflicting Values* (Oxford: Clarendon Press, 1990), chs. 7–8.

²⁰ In philosophy generally the scientizing attitude remains very strong; in fact, successor views to Quine’s naturalism are now the dominant methodological views in the discipline. The antitheoretical attitude lost its position in general philosophy decades ago and now survives almost solely in ethics.

²¹ See, e.g., Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 493–502; Thomas Hurka, *Perfectionism* (New York: Oxford University Press, 1993); and George Sher, *Beyond Neutrality: Perfectionism and Politics* (Cambridge: Cambridge University Press, 1997), ch. 9.

²² Derek Parfit, “Equality or Priority?” *The Lindley Lecture* (Lawrence, KS: University of Kansas Press, 1995); and Larry S. Temkin, *Inequality* (New York: Oxford University Press, 1993).

²³ See, e.g., George Sher, *Desert* (Princeton, N.J.: Princeton University Press, 1987); and Shelly Kagan, “Equality and Desert,” in Louis P. Pojman and Owen McLeod, eds., *What Do We Deserve? Readings on Justice and Desert* (New York: Oxford University Press, 1999), pp. 298–314.

²⁴ Shelly Kagan, “The Additive Fallacy,” *Ethics* 99 (1988): 5–31; and F. M. Kamm, *Morality/Mortality*, vol. 2: *Right, Duties, and Status* (New York: Oxford University Press, 1996), ch. 2.

attitudes have by no means disappeared. Some philosophers still reject any normative claims not derived from desires; others still look askance at all formal analysis. But the two attitudes are becoming less dominant, and as they recede, ethics is coming to theorize a wider range of views. But there is another, more indirect effect of the scientizing attitude that continues to retard moral study.

Because they are skeptical of appeals to moral intuition, scientizers are dissatisfied with what I have called structural analyses, ones that relate a moral view to principles that are more abstract but use similar concepts. To them these principles appeal to the same basic intuitions as the original view and so cannot properly justify it. There are many illustrations of this dissatisfaction.

In a recent book on desert, George Sher says that Moore's defense of retributivism in terms of his principle of organic unities is "inconclusive," since it "merely restates what the retributivist needs to explain." Though intuitive appeals like Moore's are not worthless, they are at best a "prologue" that should lead to "independent justifications of our beliefs about desert."²⁵ Similarly, Rawls allows that a pluralist view that weighs its values intuitively can describe the structure of its comparative judgments, for example, on indifference graphs, but adds that since these graphs give no "constructive criteria" establishing the judgments' reasonableness, what results is "but half a conception."²⁶ Or consider nonconsequentialism. A great contribution of recent ethics has been to show that nonconsequentialist views have a self-referential or "agent-relative" structure. When they forbid, for example, killing, they tell each person to be especially concerned that *he* does not kill, even if the result is a somewhat greater number of killings by other people.²⁷ But rather than being seen to support nonconsequentialism, by clarifying its structure, this analysis has been taken to generate objections against it. How can it be rational, some have asked, to avoid one act of killing if the result is more killings overall? If killing is bad, should we not try to minimize killing by everyone?²⁸ Here it is not taken to be a sufficient answer to point to the intuitive appeal of an agent-relative prohibition against killing or even of the abstract structure it embodies. What is demanded is a justification of agent-relativity of some deeper, more philosophical kind.

This dissatisfaction has led many contemporary philosophers to search for foundational justifications of moral views, ones that relate them to other concepts they see as somehow more secure. These justifications have taken many forms. In some the foundational principles appealed to are still moral. Thus, a prominent justification of retributivism appeals to ideas about fairness, saying it is unfair if the majority

²⁵ Sher, *Desert*, pp. 72, 19.

²⁶ Rawls, *A Theory of Justice*, pp. 39, 41.

²⁷ Bernard Williams, "A Critique of Utilitarianism," in J. J. C. Smart and Bernard Williams, *Utilitarianism For and Against* (Cambridge: Cambridge University Press, 1973), p. 99; Parfit, *Reasons and Persons*, pp. 143–48; Nagel, *The View from Nowhere*, pp. 152–56, 175–85.

²⁸ Samuel Scheffler, *The Rejection of Consequentialism: A Philosophical Examination of the Considerations Underlying Rival Moral Conceptions* (Oxford: Clarendon Press, 1982), ch. 4; and Shelly Kagan, *The Limits of Morality* (Oxford: Clarendon Press, 1989).

of people have restrained their self-interest by obeying the law but a criminal has not. He has gained the benefit of others' restraint without paying a similar cost himself, and punishing him removes that imbalance.²⁹ Here the proposed foundational claim is at a coordinate level to those being justified; fairness is not a more abstract concept than desert. But other justifications start from very abstract principles—for example, that everyone should be treated with equal respect and concern,³⁰ or that those acts are wrong that are forbidden by rules no one could reasonably reject.³¹ They then claim that the best interpretation of these principles, guided by normative views but not by ones directly favoring a given moral position, can support that position and therefore justify it philosophically. The grand exemplar of this approach is Rawls, who says the correct principles of justice are those that would be chosen by rational contractors in a specified initial position. Rawls's general contractarian claim is best understood as a moral one, and moral judgments also guide his specification of his initial position. But since these judgments do not directly concern the principles the contractors choose among, the contract provides an "Archimedean point" for justifying specific claims about justice.³²

These last analyses shade into ones that are more explicitly ambitious, claiming to derive specific moral views from the logic of moral language or the definition or purpose of morality. R. M. Hare claims that the language of morals, properly understood, allows only utilitarianism as a fundamental moral view; all other views misuse the language.³³ More vaguely, others claim that since the purpose of morality is to satisfy human wants and needs, any acceptable principles must concern wants and needs.³⁴ This seems to rule out perfectionist views of the good by definitional fiat, and certainly rules out goods not located within people's lives, such as distributions proportioned to their merits and the existence of ecosystems.

Yet another approach appeals to metaphysical facts, especially about the person, that it says an acceptable moral view must reflect. Rawls says that utilitarianism fails to take seriously "the separateness of persons," and that doing so leads to a more egalitarian view like his own;³⁵ Robert Nozick says the same facts about separateness support agent-relative prohibitions against using some as means to

²⁹ Herbert Morris, "Persons and Punishment," *The Monist* 52 (1968): 475–501; and Sher, *Desert*, ch. 5.

³⁰ Ronald Dworkin, "Liberalism," in Stuart Hampshire, ed., *Public and Private Morality* (Cambridge: Cambridge University Press, 1977), pp. 113–43.

³¹ T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998).

³² Rawls, *A Theory of Justice*, p. 261. In later work Rawls has abandoned this strong "Archimedean" claim about his contractarian argument, but still holds that connecting his liberal-egalitarian political views to contractarian ideas gives them a kind of justification they would not have if defended just by direct intuition.

³³ R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981).

³⁴ See, e.g., Philippa Foot, "Moral Beliefs," *Proceedings of the Aristotelian Society* 59 (1958–9): 83–104; G. J. Warnock, *Contemporary Moral Philosophy* (London: Macmillan, 1967), chs. 5 and 6; and Warnock, *The Object of Morality* (London: Methuen, 1971).

³⁵ Rawls, *A Theory of Justice*, pp. 27, 187.

benefit others.³⁶ In a similar vein, Samuel Scheffler argues that what he calls “the independence of the personal point of view” justifies agent-relative permissions. Apart from its deontological prohibitions, commonsense morality does not require people always to produce the most good impartially calculated. It permits them to give somewhat more weight to their own interests and therefore, within limits, to fall short of doing what is impersonally best. This permission can be justified, Scheffler argues, by noting that people do not rank outcomes only from an impersonal standpoint but also, and independently, in terms of their own preferences and values.³⁷

Yet another foundational approach tries to demonstrate the rationality of certain moral principles, so that failing to act on them is a species of irrationality. The most clear-headed such argument, that of David Gauthier, claims that accepting certain moral constraints is in each person’s self-interest, where self-interest is understood austere, as involving the satisfaction of her pre-existing desires.³⁸ A more high-minded view understands self-interest as involving an ideal of “flourishing” that may or may not be what a person actually wants. Here the justification of morality is that flourishing involves the moral virtues as essential components, and these virtues require one to act morally.³⁹ An especially grand form of argument, proposed by Alan Gewirth and Christine Korsgaard, tries to derive moral demands from the presuppositions of rational agency. The beliefs implicit in any rational action, it claims, commit one logically to certain moral principles, typically Kantian ones, so failing to act on those principles involves logical inconsistency.⁴⁰

Despite their differences, these arguments share the general project of trying to ground moral views in something more than direct intuition, through what I have called foundational analyses. Now, the difference between these and structural analyses is only one of degree, not of kind. It turns on whether the concepts an analysis uses are similar to those in the view it is analyzing, and similarity admits of degrees. There can even be analyses that straddle the structural-foundational divide, so they are neither clearly the one nor clearly the other. But a difference of degree is still a difference, and other analyses do clearly fit one model rather than the other. Thus, Rawls’s and Hare’s arguments are clearly foundational, while Moore’s and Broad’s are structural. There can also be similarities in how the two types of analysis are defended. Structural analyses cannot argue just from particular

³⁶ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 32–33.

³⁷ Scheffler, *The Rejection of Consequentialism*, ch. 3; for another version of metaphysical foundationalism see David O. Brink, “Self-Love and Altruism,” *Social Philosophy and Policy* 14, no. 1 (1997): 122–57.

³⁸ David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986). Because of its austerity, this argument cannot justify all the moral constraints most people find intuitively compelling.

³⁹ G. E. M. Anscombe, “Modern Moral Philosophy,” *Philosophy* 33 (1958): 1–19; and Rosalind Hursthouse, “Virtue Theory and Abortion,” *Philosophy and Public Affairs* 20 (1991): 223–46.

⁴⁰ Alan Gewirth, *Reason and Morality* (Chicago: University of Chicago Press, 1978); and Christine M. Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).

intuitive judgments to principles that fit them; if the principles are to explain the judgments, they must be independently plausible. Conversely, foundational analyses need not argue just from the general to the particular. They can and often do take the fact that an abstract idea coheres with and can explain particular judgments to be important evidence for it, as Rawls, for example, does. Both approaches can therefore use coherentist reasoning, treating all parts of a moral theory as justified by their relations to other parts and none as immune to revision or the source of all others' warrant.⁴¹ But even when they share this coherentism the two approaches differ in the type of moral theory they generate and, more specifically, in the type of moral explanation they give. Structural analyses assume that a concrete moral view can be explained by principles that express similar ideas in a more abstract way, using similar concepts at a higher level of generality; those who demand foundational analyses require a genuine explanation to relate the view to ideas that are different and more fundamental, because they concern some more fundamental topic.

It would be impressive if any of these foundational arguments succeeded, but in my view critical discussion has repeatedly shown that they do not. It would be going too far to state categorically that no such argument will ever succeed; each must be assessed on its merits. But time and again it turns out that to yield the specific views they are meant to justify, they must tacitly appeal to the very intuitive judgments they are meant to supplant. Let me give a few illustrations.

Arguments from metaphysical facts about the person need not violate the autonomy of ethics if they make the general moral claim that the correct normative principles for a thing must reflect that thing's nature.⁴² But in most cases the facts they start from either are not specific enough to support the particular moral views they are intended to or, if they are redescribed to do so, presuppose those views or something very close to them. For example, it is undeniable that persons are separate, in the sense that there are connections between states within a life that do not hold across lives. But many moral views reflect this fact other than the egalitarian ones it is usually taken to support. Consider a pluralist view that compares goods such as knowledge and achievement in a way that favors balance or well-roundedness within lives but not across them. It treats lives as morally significant units but has no distributive implications whatever. And even if separateness does have distributive implications, why must they be egalitarian? What is wrong in this respect with a desert principle requiring people's degrees of happiness to be proportioned to their degrees of virtue, or even with an anti-egalitarian

⁴¹ Because of this, analyses that I call foundational need not involve foundationalism in the sense used in epistemology, where it connotes the view that some beliefs are self-justifying and the source of justification of all other beliefs. Analyses that appeal to the language of morals or defend the rationality of morality involve something like this epistemic foundationalism, but others such as Rawls's do not. The latter are foundationalist only in my sense, which concerns explanation rather than justification.

⁴² Rawls, *A Theory of Justice*, p. 29.

principle like Nietzsche's that directs everyone to promote the excellence of the few most excellent individuals? Of course, one can redescribe the separateness of persons so it does support only egalitarian views, for example, by saying that unit gains in less happy lives count for more than similar gains in happier ones. But then egalitarianism is being built into the supposedly metaphysical facts, which no longer give it independent support. A similar point applies to Scheffler's justification of agent-relative permissions. The fact that people have a personal ranking of outcomes that can differ from the impersonal one seems perfectly well captured by a view that combines each person's personal ranking with the impersonal one, say by averaging them, and then requires him to maximize the resulting combined value. To support permissions in particular over this alternative, the metaphysical facts must be redescribed so they favor giving each person with an independent point of view the further independence to decide which point of view he will act from on particular occasions. But then the redescription again assumes something very close to the conclusion it is meant to justify.⁴³

A similar difficulty faces arguments from abstract moral principles such as ones about equal respect or rational agreement. In some cases these principles will be widely accepted, but only because they are open to different interpretations that support very different substantive views; in other cases they do favor one view, but have contentious features that cannot be independently justified.⁴⁴ The former is true of arguments from equal respect. Their foundational principles are unexceptionable if they say only that a moral view must give all people equal standing, but this claim has no substantive implications. In particular, it does not imply that people should have, even initially, equal resources or happiness; it is perfectly satisfied by views that give people equal rights to acquire resources (which they may then exercise very unequally) or say they should receive different rewards according to their different deserts. The abstract claim does not even distinguish between consequentialist and deontological views, each of which interprets respect in its own way.⁴⁵ The contrary problem arises at many points in Rawls's contractarian argument. Its basic design excludes perfectionist, entitlement, and many other views—so it is hardly a neutral device—and even views it initially allows, such as utilitarianism, are eventually excluded by stipulations that have no persuasive contractarian rationale.

⁴³ In addition, most appeals to a metaphysical fact assume a certain valuation of it; thus, Scheffler assumes that the independence of the personal point of view is not simply regrettable, as strict impartialists might hold, but has positive worth. This is another point where these appeals come close to assuming what they are meant to explain.

⁴⁴ I think most philosophers would now agree that a similar dilemma faces arguments from the language of morals or the definition of morality. Most would say its having substantive moral implications actually counts against a view on these topics; an adequate view should allow the disagreements that obviously exist by remaining neutral on substantive moral questions.

⁴⁵ Consequentialist views say that respect for a person involves promoting her good, deontological views that it involves, more centrally, not sacrificing her good. This is a fundamental difference, but cannot be resolved by reflecting on the true meaning of "respect."

Or consider the suggestion that agent-relative prohibitions are justified because, by making persons “inviolable,” they give them higher moral status than they would otherwise have.⁴⁶ It may be true that these prohibitions give each person higher status by denying that she may legitimately be sacrificed whenever this will give greater benefits to other people. But they also give her lower status by denying that other people may be sacrificed to give benefits to her, and there is no independent way of saying whether the gain or loss here is greater.⁴⁷

These abstract arguments share a further difficulty with some that try to ground a moral view in another of a coordinate degree of abstractness, such as the fairness justification of retributivism. These latter arguments are difficult to generalize about and may sometimes prove illuminating. But consider the claim of the fairness justification that punishment removes an imbalance whereby some people have benefited from others’ restraint without paying a similar cost themselves. If cost is understood in a standard way, say in terms of pleasure and pain, this claim does not yield the right results about particular punishments: many crimes that intuitively call for severe punishments, such as murder, are not ones people pay a high cost in avoiding, since they are not ones they are strongly tempted to commit. So the justification must revise its understanding of cost, and in doing so make some finely balanced claims about the relative positions of criminals and law-abiding citizens. And we can then ask whether those claims really are explanatory of retributivism. On one side is a simple judgment many find intuitively compelling, about the intrinsic appropriateness of punishing the guilty; on the other are some precarious claims about the balance of costs and benefits. Is it really an advance to relate the former to the latter? I am not suggesting that everyone must accept retributivism; many may on reflection reject it. But their reflection will be most productive if it considers retributivism in its own right rather than detouring through a view that seems too convoluted to underlie it.⁴⁸ A similar point applies to more abstract arguments. Rawls tries to ground his egalitarian claim that it is most important to improve the condition of the worst off by arguing that his contractors will not know the probabilities of their occupying different positions in society and, lacking that knowledge, will prefer those principles whose worst outcome for them is least bad. Critics have questioned whether either of these stipulations is reasonable—in my view, persuasively. But even if they are reasonable, how much is gained by relating an egalitarianism many find compelling in itself to these finely balanced claims about rational choice?

It was not with pleasure that the early twentieth-century theorists said, as they often did, that the resolution of a fundamental moral issue comes down to each

⁴⁶ Kamm, *Morality/Mortality*, vol. 2, ch. 10.

⁴⁷ Kagan, “Replies to My Critics,” *Philosophy and Phenomenological Research* 51 (1991): 919–20.

⁴⁸ In his version of the fairness justification Sher equates the cost a criminal avoids with the strength of the moral prohibition he violates (*Desert*, pp. 81–2), but this brings the justification very close to the Moorean formulation it is meant to improve upon.

person's intuitive judgment about the question when it is clearly framed. They would have been delighted if more satisfying justifications were available. But if they are not available, the widespread search for foundational analyses in recent ethics has been quixotic. This is not a decisive objection; many philosophical projects are quixotic, and the discipline still learns from their failure. But in this case the pursuit of grand justifications has in several ways skewed the field's priorities.

First, it has helped keep the focus of normative ethics narrow; as well as rejecting moral views for being extravagant, philosophers can dismiss them for lacking deep foundations. In some cases this dismissal has not had such chilling effects. Thus, the charge of some that agent-relative prohibitions are irrational has not stopped others from exploring their details. But, in another case, the novel entitlement theory of justice described by Robert Nozick has received less constructive discussion than it deserves in part because of the complaint that Nozick's treatment leaves it "without foundations."⁴⁹ This might be a reasonable complaint if competing views did have such foundations; it is not when they do not.

Second, the search for foundations has diverted philosophers' attention from the details of moral views. Faced with a set of ideas about value or duty, a theorist can go in either of two directions. She can look into the view, to discover its structure and the way it generates specific claims, or she can try to connect it to larger ideas about, say, the nature of morality or the status of persons as free and rational. And the more she does of the latter the less she will do of the former. In fact, a climate of foundational inquiry can lead even theorists sympathetic to a given view to spend more time answering others' abstract objections to it than examining its specific contents. But in ethics as in many fields, God (or the devil) dwells in the details. Our commonsense ideas about any moral topic are at best loosely organized and demand systematization. Sometimes this task cannot be completed: the ideas turn out to conflict, and we can only retain some by abandoning others. But at other times theorizing a view can show it has hidden strengths. For example, what seemed at first a loose collection of judgments can turn out to have a sophisticated underlying structure, even a mathematical one, that everyday moralists follow though they are unaware of doing so. Or the view can have a surprising unity. It may include, say, a pair of plausible intuitive judgments that look independent, so we can wonder whether they are even consistent. But then analysis reveals that, perhaps given a third plausible judgment, the two entail each other, so the view has an impressive internal integrity. Or the view can be surprisingly related to several others, all of which turn out to instantiate the same more abstract idea. Or, examined closely, it can generate fascinating new questions and offer intuitively satisfying answers to them—sometimes just one answer, sometimes a set of competing ones. Any of these developments counts in favor of a view's acceptability, as internal contradictions count against it. But none will

⁴⁹ Nagel, "Libertarianism without Foundations," *Yale Law Journal* 85 (1975): 136–49.

emerge if philosophers are busy batting around abstract conditions and chasing foundational dreams.

In fact, the search for moral foundations can lead to high-minded distortions of the moral phenomena. Consider the increasingly popular view that most desires are for objects thought of as good or as somehow supported by reasons. It has led one writer to say that thirst involves a recognition that we have reason to drink,⁵⁰ but surely no one actually gets thirsty in this highly intellectualized way. As Plato said, what we want when we are thirsty is not a good drink, but just a drink.⁵¹ And even Plato was too restrictive in limiting this type of desire to the bodily appetites. Anything we can desire on the basis of an evaluative judgment we can also desire apart from such judgments; as well as pursuing knowledge because we believe it is intrinsically good, we can pursue it from simple curiosity. Moore, Rashdall, and Ross recognized this, holding that alongside forms of virtue directed at the right and the good as such are forms that involve no evaluative thoughts, such as simple compassion.⁵²

Other distortions are more disturbing. A persistent temptation for foundational views, especially ones that affirm the rationality of morality, is to ground moral requirements in facts about the self: in how acting on them promotes the agent's self-interest or flourishing or relates to a "practical identity" or conception she has of herself as a specific type of person.⁵³ But if they take this line, these views imply that my ultimate reason for benefiting another person—say, by relieving his pain—concerns how this will affect my life. And this is not the right ultimate reason, which concerns how my action will affect the other's life: my fundamental reason to help him is to help him. In a trivial sense morality has to concern me, since it has to tell me how to act. But it does not follow that the ground of its requirements must always be located in me, and views that place it always there are in that respect distortions. This is reflected in the fact that people who were motivated by what these views say are their ultimate reasons—people whose ultimate motive for helping another was to promote their own flourishing or to live up to their own practical identity—would be acting in an objectionably narcissistic way. It is not admirable but morally self-indulgent to relieve another's pain primarily from concern to be a virtuous reliever of pain.⁵⁴ Now, a moral view can say that people will only act successfully on their ultimate reasons if they do not try consciously to do so. Thus, consequentialist views often say that people will only produce the best consequences if they do not aim at those consequences directly but instead follow

⁵⁰ Scanlon, *What We Owe to Each Other*, p. 38.

⁵¹ Plato, *Republic*, trans. G. M. A. Grube (Indianapolis: Hackett, 1974), 438a–39b.

⁵² Moore, *Principia Ethica*, 179; Hastings Rashdall, *The Theory of Good and Evil* (London: Oxford University Press, 1907), vol. 1, pp. 119–24; Ross, *The Right and the Good*, pp. 160–63, 170–73.

⁵³ For the latter view, see Korsgaard, *The Sources of Normativity*.

⁵⁴ On moral self-indulgence see Bernard Williams, "Utilitarianism and Moral Self-Indulgence," in his *Moral Luck: Philosophical Papers 1973–80* (Cambridge: Cambridge University Press, 1981), pp. 40–53.

certain simple rules. But it would be very odd for the foundational views I am discussing to make this type of claim. That would involve their saying that action on what they say are our ultimate reasons is objectionable in itself, and not just because of its effects, and surely a moral view cannot so directly condemn its own acceptance. That is a sign that something has gone deeply wrong—that the search for philosophical foundations has led theorists to place the source of our reasons where it intuitively cannot be.

It is in fact striking how many strands in contemporary ethics, of Kantian, Aristotelian, and other inspirations, approach the subject in an essentially self-centered way. They see the fundamental task of the moral agent as being to reflect on her own desires, or care about her own flourishing or integrity, or give moral laws to herself and so achieve her own autonomy. On all these views the agent's fundamental orientation is toward herself. There are, of course, exceptions; Thomas Nagel says very clearly that another person's pain by itself gives me reason to relieve it.⁵⁵ But a great deal of currently influential writing seems to assume that moral reasons must be not only trivially but also substantively about the self. The causes of this are partly philosophical: disbelief that there could be underivative other-regarding reasons and distrust of the intuitions that say there are. But there may also be social causes. May the self-centeredness of so much contemporary ethical theorizing be the predictable product of a narcissistic age, offering moral foundations for the Me Generation?

Finally, there is an issue of philosophical style. The early twentieth-century theorists wrote about ethics simply, directly, and even (one thinks, for example, of Rashdall) with wit. But much contemporary ethical writing has a tone of high intellectual earnestness. Sidgwick thought the progress of ethics had been impeded by the desire to edify; today the greater danger is the desire to be philosophically profound. At the very least, the pursuit of grand foundational projects does not encourage an engaging lightness of touch.

Any assessment of the current state of ethics must address the mixed influence of Rawls. *A Theory of Justice* is a great book, with many novel ideas and an original proposal for uniting them. And in a lecture given shortly after that book's publication, Rawls made methodological recommendations similar to those of this essay: that at least for now philosophers should set aside questions of moral truth and examine the structures of the moral views people actually hold.⁵⁶ But these recommendations have had much less influence than Rawls's practice or what was taken to be his practice in *A Theory of Justice*. Many even of those who do not think that book's contractarian argument succeeds think it is the right type of argument, aiming to justify a moral view not by direct intuitions but by something more

⁵⁵ Nagel, *The View from Nowhere*, pp. 156–62.

⁵⁶ Rawls, "The Independence of Moral Theory," *Proceedings of the American Philosophical Association* 48 (1974–5): 5–22. But Rawls includes in the "structure" of a moral view an associated "conception of the person," which sounds more foundational than the more purely mathematical analysis of structure I have been urging.

philosophically ambitious. And Rawls's later work has reinforced this tendency, with its emphasis on grand abstractions such as persons' "higher-order interest" in forming, revising, and pursuing a conception of the good and the ideal of "public reason." Rawls's writing in moral philosophy has had many effects, but one has been to encourage a style of argumentation that is more high-flown than it is productive.

How, then, should normative ethics proceed? I take it that on most views its starting-point is the moral judgments people actually make. And it is worth emphasizing that people do make these judgments. Despite the claims of some cultural analysts, categorical moral evaluations are not disappearing in the face of moral relativism. No one says that rape, racism, and genocide are morally all right for you if you believe they are all right; everyone simply condemns them as wrong. (Though some describe themselves as relativists, their moral practice repeatedly belies that claim.) And though the content of these judgments may change somewhat over time, their core remains the same. There is less concern today than fifty or a hundred years ago with the ethics of consensual sexual activity, and more with discrimination and intolerance. But the central moral prohibitions against coercion and willful harm remain essentially unchanged, as do the central recommendations of compassion and fidelity. There are metaethical questions about how these prohibitions and recommendations are to be understood: in a realist way, as making claims that can be true or false, or as merely expressing attitudes. But these questions do not affect, except at the margins, either the content of these judgments or their importance. Whatever metaethics decides, the practice of making moral judgments will continue essentially as before.

In their initial form, however, these judgments require philosophical analysis, both to understand and to assess them. In my view this analysis is best done by addressing them in their own right, by seeking to identify more abstract principles that, while using similar concepts, can organize and explain them, and by investigating their details. In fact, these two tasks often go together, since trying to describe a view's structure comprehensively can force one to address specific issues one might otherwise not have thought of. Much valuable work of this kind has been done in recent ethics. The clear description of agent-relative prohibitions and permissions, which require us to care especially that *we* do not kill and allow us to give more weight to *our* interests, is, when separated from bogus demands for deep justifications, an immense contribution to our understanding of commonsense and deontological views. Other writers have shown how, in formulating their prohibitions, these views can use either or both of the distinctions between harms one causes and those one merely allows, and between harms one intends as an end or means and those one merely foresees; they have also discussed how these two distinctions might combine with each other. This is not a topic where a completely satisfactory analysis has yet emerged. Perhaps that analysis requires a more subtle application of the principle of organic unities than has yet been considered; perhaps it is simply not possible. Even so, the identification of the two distinctions has

greatly improved our understanding of this family of views.⁵⁷ There have also been subtle suggestions about how a deontological view can weigh its prohibitions against the overall good an action will cause without aggregating that good in a simple additive way. Thus, the view can make it permissible to kill one innocent person to save some large number of other people from being killed or otherwise seriously harmed, but not to save any number of people from mild headaches.⁵⁸

There have been similar insights into views about the good. Thus, it has been shown that these views need not require everyone to maximize the good. They can instead require agents only to aim at outcomes that in one of several senses are “good enough”; this identification of “satisficing” consequentialism has again widened the range of moral views philosophers can consider.⁵⁹ There have also been discussions about whether the onerousness of one person’s duty to pursue good outcomes should be affected by other people’s failure to fulfill their duty, and if not, why not.⁶⁰ And there have been illuminating analyses of particular values, such as a distinction between two kinds of broadly egalitarian view⁶¹ and a searching examination of the complexities of one of them,⁶² as well as the beginning of a sophisticated structural analysis of judgments about desert.⁶³ There remain many moral views that have not received the structural analysis they deserve, but on others significant progress has been made.

In some cases this analysis will contribute significantly to assessing the views. Thus, it may be that while one of two competing views resists structural analysis, because it contains ineliminably contradictory elements, another can be beautifully unified by an attractive general principle. This result supports the second view against the first. But in other cases even complete analysis may leave us with difficult judgments to make. If each of two views can be successfully unified, the choice between them comes down to an overall intuitive assessment that we may find difficult to make and that different people may make differently. Here there may be neither

⁵⁷ See, e.g., Philippa Foot, “The Problem of Abortion and the Doctrine of Double Effect,” in her *Virtues and Vices* (Berkeley, CA: University of California Press, 1973), pp. 19–32; Judith Jarvis Thomson, “Killing, Letting Die, and the Trolley Problem,” *The Monist* 59 (1976): 204–17; Warren S. Quinn, “Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing,” *Philosophical Review* 98 (1989): 287–312; Quinn, “Actions, Intentions, and Consequences: The Doctrine of Double Effect,” *Philosophy and Public Affairs* 18 (1989): 334–51; Jeff McMahan, “Killing, Letting Die, and Withdrawing Aid,” *Ethics* 103 (1993): 250–79; Kamm, *Morality/Mortality*, vol. 2, chs. 1–7.

⁵⁸ Samantha Brennan, “Thresholds for Rights,” *Southern Journal of Philosophy* 33 (1995): 143–68; Kamm, *Morality/Mortality*, vol. 1: *Death and Whom to Save from It* (New York: Oxford University Press, 1993), chs. 8–10; and Scanlon, *What We Owe to Each Other*, 229–41.

⁵⁹ Michael Slote, *Common-Sense Morality and Consequentialism* (London: Routledge & Kegan Paul, 1985), ch. 3; and *Beyond Optimizing: A Study of Rational Choice* (Cambridge, Mass.: Harvard University Press, 1989), chs. 1–2.

⁶⁰ Liam B. Murphy, *Moral Demands in Nonideal Theory* (New York: Oxford University Press, 2000); Brad Hooker, *Ideal Code/Real World* (Oxford: Clarendon Press, 2000).

⁶¹ Parfit, “Equality or Priority?”

⁶² Temkin, *Inequality*.

⁶³ Kagan, “Equality and Desert.”

intuitive certainty for one person nor agreement among persons. This may also be the result in the more likely case where each view can be unified to a considerable degree but not entirely, so there is a plausible general principle that captures many of the intuitive judgments associated with it but requires others to be abandoned. But this is just the situation we face if, as I have argued, there are no magic-bullet arguments in normative ethics and intuitive judgments of some kind are all we have to go on. We can play such judgments off against each other, including judgments at different levels of generality. If we think some judgments are more likely to be distorted we can give them less weight, though if we do the correction will always come from other intuitive judgments. And if we think some judgments are especially reliable, we can place more trust in them. This methodology leaves philosophical analysis only some comparatively modest tasks: to clarify the issues our judgments must address and to help us understand the views that resolve them differently. But those are not insignificant tasks; on the contrary, and unlike the high flights of foundational analysis, they can actually yield results.

The main themes of this chapter were anticipated by Friedrich Nietzsche. In *Beyond Good and Evil* he urged moral philosophers to own up

to what alone is justified so far: to collect material, to conceptualize and arrange a vast realm of subtle feelings of value and differences of value which are alive, grow, beget and perish . . . all to prepare a *typology* of morals.

To be sure, so far no one has been so modest. With a stiff seriousness that inspires laughter, all our philosophers demanded something far more exalted, presumptuous, and solemn from themselves as soon as they approached the study of morality: they wanted to supply a *rational foundation* for morality—and every philosopher so far has believed that he has provided such a foundation. How remote from their clumsy pride was that task which they considered insignificant and left in dust and must—the task of description—although the subtlest fingers and senses can scarcely be subtle enough for it.⁶⁴

Subtle fingers and senses, yes: normative ethics will do best if it does more of the work Nietzsche recommends and less of the kind that would have him rolling on the floor.

⁶⁴ Friedrich Nietzsche, *Beyond Good and Evil*, trans. Walter Kaufmann (New York: Vintage, 1966), sec. 186.

PART II

Comparing and Combining Goods

This page intentionally left blank

Value and Population Size

Just because an angel is better than a stone, it does not follow that two angels are better than one angel and one stone. So said Aquinas (*Summa contra Gentiles* III, 71), and the sentiment was echoed by Leibniz. In section 118 of the *Theodicy* he wrote: “No substance is either absolutely precious or absolutely contemptible in the sight of God. It is certain that God attaches more importance to a man than to a lion, but I do not know that we can be sure that he prefers one man to an entire species of lions.” Even Kant was bitten by this bug. In one of his pre-Critical works he was moved to say, *à propos* of lice, that even though they “may in our eyes be as worthless as you like, nevertheless it is of more consequence to Nature to conserve this species as a whole than to conserve a small number of members of a superior species.”¹

In these passages Aquinas, Leibniz, and Kant gave expression to a distinctive and interesting view about the value of animal species and animal populations. At its simplest, this is the view that there is a special value in the existence of animal species or in the existence of a wide variety of different animal species. But the view also goes deeper than this. An animal species, after all, is nothing over and above the individual animals that make it up, and the value that it contributes to the world must therefore be some function of the values contributed by those individual animals. At the deepest level, what the view expressed by Aquinas,

In the course of writing this chapter I have benefited from the comments of John Leslie, Dennis McKerlie, John Heintz, J. J. MacIntosh, and two referees. Robert Woodrow and Verena Huber-Dyson helped me with some of the mathematics.

¹ Immanuel Kant, *Allgemeine Naturgeschichte und Theorie des Himmels* (1755), in *Kants gesammelte Schriften*, 24 vols. (Berlin: Königlich Preussischen Akademie der Wissenschaften, 1902–66), vol. 1, p. 354. The quotations from Aquinas, Leibniz, and Kant are drawn from A. O. Lovejoy, *The Great Chain of Being* (Cambridge, Mass.: Harvard University Press, 1936), pp. 75, 225, 362.

Leibniz, and Kant holds is that the value that an individual animal contributes to the world is not constant but varies with the number of other animals in his species. When the number of other animals in his species is small, his own existing contributes a fairly large amount of value to the world, and his passing out of existence deprives it of a fairly large amount of value as well. But when the number of other animals in his species is large, the issue of his existing or not existing is not nearly so significant.

Let us call the view expressed by Aquinas, Leibniz, and Kant the “variable value view”; I think this view is both an attractive one and one that many of us already adopt in our thinking about animal populations. Consider for instance our attitude to an endangered species like the whooping crane. If there were a large number of whooping cranes in the world, then I am sure none of us would think that the birth of another whooping crane was an event of any great significance or that the death of a whooping crane was something that it was worth spending any large sum of money to prevent. But things are different when the whooping crane population falls as low as sixteen. Then individual whooping crane births are widely reported in the international press, and large sums of money are spent both to encourage the births of new whooping cranes and to prevent the deaths of any existing whooping cranes. When we reflect that this same money could easily be spent on medical or safety facilities for humans and that if it was it could easily save a number of human lives, it is not at all absurd to suggest that, when the whooping crane population gets to be small enough, the value of an individual whooping crane life increases in our eyes to a point where it is not all that much smaller than the value of an individual human life.

Although Aquinas, Leibniz, and Kant applied the variable value view primarily to animal populations, the same view can also be applied to human populations. It will then hold that the value that an individual human being contributes to the world is not constant but varies with the number of other human beings in the world, being much larger when that number is small than it is when that number is large. This second application of the variable value view is also an attractive one, and I think it is also one that many of us already make in our thinking about human population increase. When we imagine situations in which the human population has been reduced to a mere handful of people, as it is in the biblical story of Noah, or as it might be in the event of a nuclear holocaust, one of the first things that springs to our minds is the tremendous importance of increasing that population and of starting to repopulate the world with humans. One of the principal duties of the handful of people who survived a nuclear holocaust would be to procreate, and we think this duty would be binding on them even if, as a result of limited supplies and resources, procreating would diminish their own well-being considerably. But when the human population gets to be as large as it is now, with some three or four billion people in the world, further increases in its size do not seem nearly so important to us. If it were possible to add to the present population of the world another five million people who would be just as well-off

as existing people, and to do so in a way that did not make any other people worse off, then I think many of us would regard the addition of these five million people as a good thing. But we would surely not regard it as a *very* good thing; and we would surely not think that any *very* serious wrong had been done if the addition was not made.

The variable value view is not captured by either of the two consequentialist principles that are most commonly applied to questions about human population increase—namely, the average principle and the total principle—and this does a lot to explain why these two principles are so unattractive. The average principle is unattractive primarily because of its consequences at low levels of population. These are the levels at which we think population increases have the most value, yet the average principle gives them no value if they are not accompanied by an increase in the average well-being per person and a negative value if they are accompanied by even the smallest decrease in the average well-being per person.² The average principle's consequences at high levels are not nearly so unattractive, for, as we have seen, we do not think that population increases at these levels are nearly so important. But the consequences are still not entirely attractive. I think most of us believe that population increases at high levels have at least a little value, and if we do, we will have to regard a principle that gives them no value as one that goes a little too far. Unlike the average principle, the total principle has its most unattractive consequences at high levels of population. At these levels it gives far too much weight to population increases, and it continues to require these increases far beyond the point where most of us think they have ceased to be morally important. This can be brought out most strikingly by showing that the total principle is committed to accepting what Derek Parfit (perhaps following McTaggart)³ has called the “repugnant conclusion.” Let us try to imagine an ideal world, one in which there is an exceptionally large number of people at an exceptionally high level of well-being. Then if the total principle is correct there is another world that is better than this one, a world in which there is some much larger number of people at a much lower level of well-being, a level, in fact, at which their lives contain a barely positive amount of well-being and are therefore “barely worth living.” This conclusion is, I think, rightly called repugnant, but a principle that gives as much value to population increases at high levels as the total principle does cannot avoid accepting it. The total principle's consequences at low levels of population are much more attractive, for at these levels the principle gives value to many population increases that involve a decrease, and even a significant decrease, in the

² Here as elsewhere in this paper I use “well-being” to refer neutrally to whatever state of human beings a consequentialist principle takes the value they contribute to the world to be a function of. For a utilitarian principle this state will be happiness, for a perfectionist principle some state of human perfection or excellence, and so on. The value human beings contribute to the world need not, of course, be a function only of their well-being. If that were the case, the variable value view would be ruled out from the start.

³ See J. M. E. McTaggart, *The Nature of Existence* (Cambridge: Cambridge University Press, 1927), vol. 2, pp. 452–53.

average well-being per person. But it is once again open to question whether these consequences are entirely attractive. We have seen that population questions would take on a special urgency in the event of a nuclear holocaust, and it is not clear that a simple summative principle like the total principle can account for this special urgency. It is of course true that when the population level gets very low the survival of the human race is threatened, and that since the survival of the human race is a necessary condition of there being any value in the future the total principle will tell us to do a fair amount to ensure it. The trouble is that in this case a fair amount is not enough. Imagine that an all-powerful being (I hope I am not begging any questions when I call him the devil) offers us a wager that he says we have a .51 chance of winning and a .49 chance of losing. If we win he will make it the case that at any time in the future when there would otherwise have existed a certain number of people at a certain average well-being per person there will exist twice that number of people at the same average well-being; whereas if we lose he will cause the human race to die out. I think most of us would say we ought to reject this wager of the devil's; but the total principle says we ought to accept it.⁴

It is important to be clear about the exact nature of these difficulties for the average and total principles. I do not want to deny that the average and total principles will require many population increases at low levels of population and forbid many increases at high levels. They will require increases at low levels if these allow for a more extensive division of labor and thus greater economic productivity, leading to a greater average well-being per person (and also total well-being). And they will forbid increases at high levels if these place too great a strain on scarce resources, leading to a smaller total well-being (and also average well-being). In these cases the average and total principles will make judgments based on the *side effects* that increases in the human population will have on already existing people. The difficulties I have raised concern in the first place cases where these side effects do not obtain. I have argued that, even when population increases will have no effect on the average well-being, we think they are more important at low levels than at high levels, and that the average and total principles cannot capture this view. But the difficulties also extend to cases where the side effects do obtain. If the average principle gives too little weight to population increases at low levels when they do not affect the average well-being, it will also give too little weight to

⁴ An adherent of the total principle may try several responses to this argument. He may say, first, that our refusal to accept the devil's wager does not show anything about our attitude to values, but only that we are extremely risk averse when the stakes are very high. But this seems to me perverse. Why attribute an irrational attitude to risk to ourselves when we can explain the same phenomena by simply allowing that we might have something other than the most simpleminded approach possible to the quantification of values? An adherent of the total principle may also say that we reject the devil's wager because we imagine that without it there may be an infinite number of human beings in the future, and twice infinity is still infinity. This response will not do, for the total principle can never consider the possibility of infinite future amounts of well-being. If it does, it will have to allow that any actions that have even the remotest probability of being followed by an infinite future amount of well-being—that is, just about any actions at all—have the same (infinite) expected value and are therefore all morally indifferent.

these increases when they do affect the average well-being; and the opposite holds (at high levels) for the total principle. Some adherents of the total principle have claimed that the human population is already so large that further increases in its size will lower the total well-being in the world, so that the total principle no longer requires us to make these increases. I think these claims are fanciful and that on any plausible assumptions the total principle will be requiring population increases for a good while to come. If this is the case, however, the total principle has unattractive consequences for choices we are making in the actual world today, with all possible side effects taken into account.

There are several consequentialist principles that do capture the variable value view as it applies to human populations, but in this paper I want to examine the two that I think are the most plausible. To do this I need to introduce a family of graphs, which I will call “population value graphs.” The vertical axis on one of these graphs measures the value that a human population contributes to the world (at a given time), while the horizontal axis measures the size of the population (at that time), and lines across the graph, or population value lines, show how the value that a population contributes to the world (at a given time) is a function of its size, given a fixed average well-being per

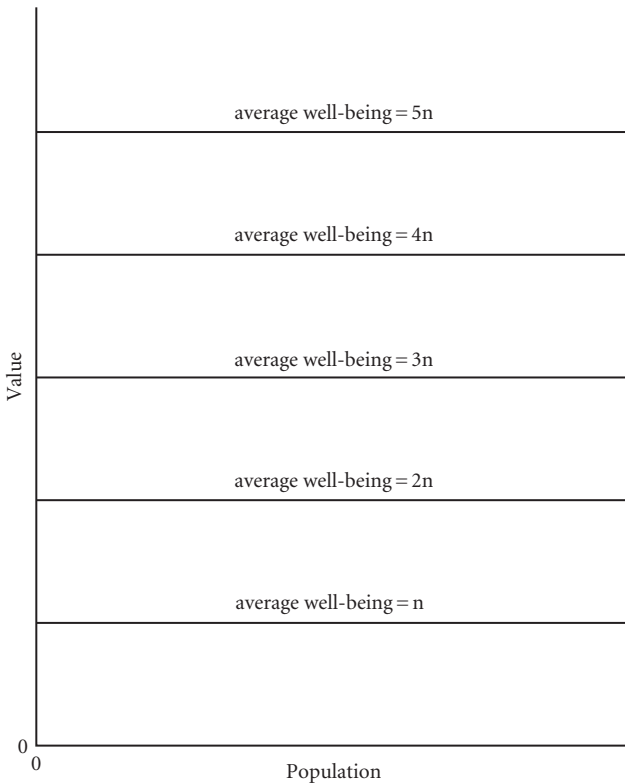


FIGURE 2.1 *The Average Principle*

person. Different consequentialist principles calculate the values measured on the vertical axis in different ways, but for simplicity's sake I will assume that they all then tell us to maximize the sum of these values across times. It may help to explain the operation of these graphs if we show how the two most familiar principles—namely, the average and total principles—can be represented on them. A population value graph for the average principle contains a family of evenly spaced population value lines that are both straight and horizontal (Fig. 2.1). Since the average principle holds that the value that a population contributes to the world is independent of its size, a move to the right along one of these lines (a move that represents an increase in the size of the population without any change in the average well-being) is not accompanied by any vertical rise (which represents an increase in value), and the only way to achieve such a rise is to move to a line that represents a higher average well-being. The lines on a graph for the total principle, by contrast, do rise to the right (Fig. 2.2). They are straight lines that begin at the origin and rise to the right, with the lines representing a higher average rising more steeply than those representing a lower average.

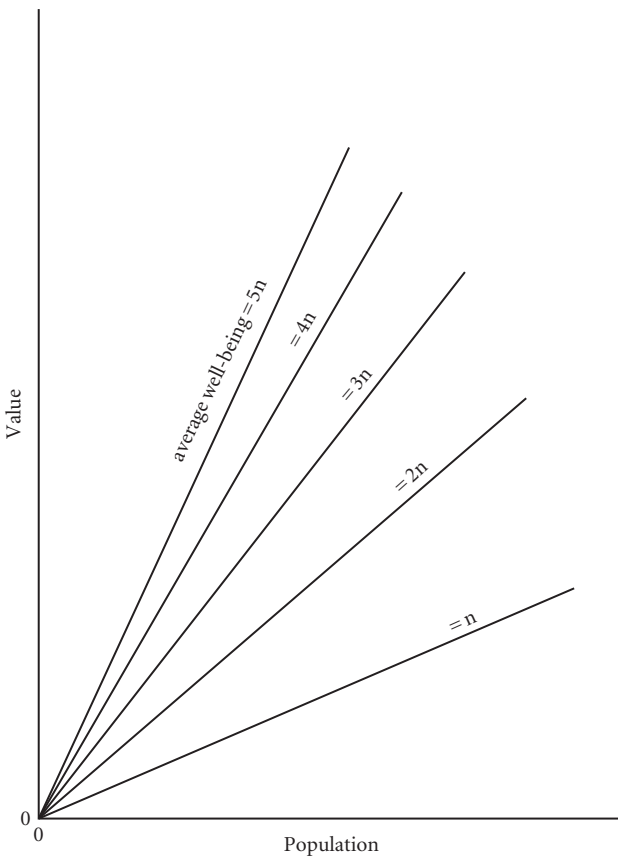


FIGURE 2.2 *The Total Principle*

Now the fact that the average and total principles do not capture the variable value view is reflected in the straightness of the population value lines in Figures 2.1 and 2.2. The way to capture the view is to give up the assumption that these lines have to be straight and to consider instead the possibility that they might be curves, in particular, that they might be curves that begin by rising quite steeply from the origin and then flatten out as they move to the right. One attractive such possibility is represented in Figure 2.3, in which every population value line begins by rising steeply from the origin (much more steeply, it should be noted, than the corresponding line on the graph for the total principle) and then flattens out to converge on a horizontal straight line, a line to which it is therefore asymptotic.

Because of the way in which it captures the variable value view, the principle represented in Figure 2.3 (let us call it V_1) avoids many of the unattractive consequences of both the average and total principles. At low levels of population V_1 gives more value to population increases than either of these principles, and it therefore does more to

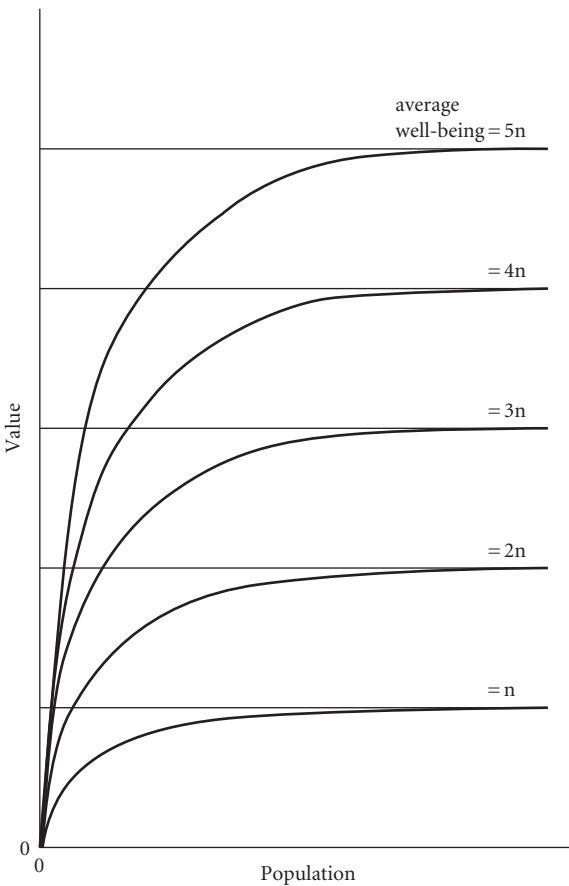


FIGURE 2.3 *Principle V_1*

capture the urgency that would attach to population questions in the event of a nuclear holocaust. At the same time, V_1 tells us to reject the devil's wager. Since the population value lines it generates flatten continuously as they rise to the right, V_1 holds that a doubling of the population size that is not accompanied by any change in the average well-being never does as much as double the value contributed by the population, and it therefore holds that we should not accept the devil's wager unless the chances of our winning it are made somewhat better. If the population is not going to rise above a handful of people in the future, then our chances will not have to be made very much better; but if it is going to be very large in the future, say as large as it is now, then they will have to be made very much better indeed. The principle V_1 also avoids the unattractive consequences that the average and total principles have at high levels of population. In particular, while holding with the total principle that of any two populations at the same average well-being the larger one is always the better one, V_1 avoids the repugnant conclusion. Since the population value lines in Figure 2.3 are asymptotic to horizontal lines, V_1 says there is an upper limit on the value that a population at a given average well-being can contribute. If the "ideal" world we imagine has more value than the upper limit for worlds in which everyone's life is "barely worth living," as it is surely certain to do, then V_1 says we cannot improve on this world by moving to another in which everyone's life is "barely worth living," no matter how many people that second world contains.⁵

Although V_1 accepts the variable value view as it applies to population size, it does not accept any similar view about the average well-being but holds with both the average and total principles that, given a constant population size, the value of a fixed increase in the average well-being is itself always constant, so that a doubling of the average well-being that is not accompanied by any change in the size always doubles the value contributed by the population. This is reflected in the even spacing of the asymptotes in Figure 2.3, which ensures that whenever we move from a point on a line representing one average well-being to the point directly above it on the line representing twice that average well-being, we always arrive at a point twice as far up the graph as the one from which we started. But a consequentialist principle can also apply something like the variable value view to the average well-being and make the value of a fixed increase in the average get smaller as the average gets higher. A principle that does this is represented in Figure 2.4, in which the asymptotes are no longer evenly spaced, but draw closer together the farther we move up the graph.

The principle represented in Figure 2.4 (let us call it V_2) has similar consequences for questions about population increase as V_1 , but there are at least two respects in which I think it is more attractive than V_1 (for a third, see below). As we have seen, V_1 holds that a doubling of the population size that is not accompanied by any

⁵ Consider a principle that is like V_1 except that it generates population-value lines that are asymptotic to lines that rise slightly to the right. Although this principle gives much less weight to population increases at high levels than the total principle does, it is still committed to accepting the repugnant conclusion. This is sufficient, I think, to make it a less attractive principle than V_1 .

change in the average well-being never does as much as double the value contributed by the population; but, at the same time, it holds that a doubling of the average that is not accompanied by any change in the size always doubles that value. This means that, given a choice between (merely) doubling the average and (merely) doubling the size, V_1 always tells us to (merely) double the average. If we attach great value to population increases at low levels, we should find this an unattractive consequence and should take it as a reason for preferring a principle like V_2 , which sometimes tells us to (merely) double the size. The second respect in which V_2 is more attractive than V_1 concerns another wager of the devil's, which we once again have a .51 chance of winning and a .49 chance of losing. If we win, the devil will make it the case that, at any time in the future when there would otherwise have existed a certain number of people at a certain average well-being, there will exist the same number of people at twice that average well-being; whereas if we lose, he will cause the human race to die out. If we think we ought to reject this wager, as I think many of us will, we should hold it against V_1 that, like the average and total principles, it tells us to accept it. The principle V_2 , however, tells us to reject the wager.

Whatever the differences between V_1 and V_2 , they are less important than the fact that they both capture the variable value view, and both avoid many of the unattractive consequences of the average and total principles. However, V_1 and V_2 are open to at least two objections that are worthy of examination.

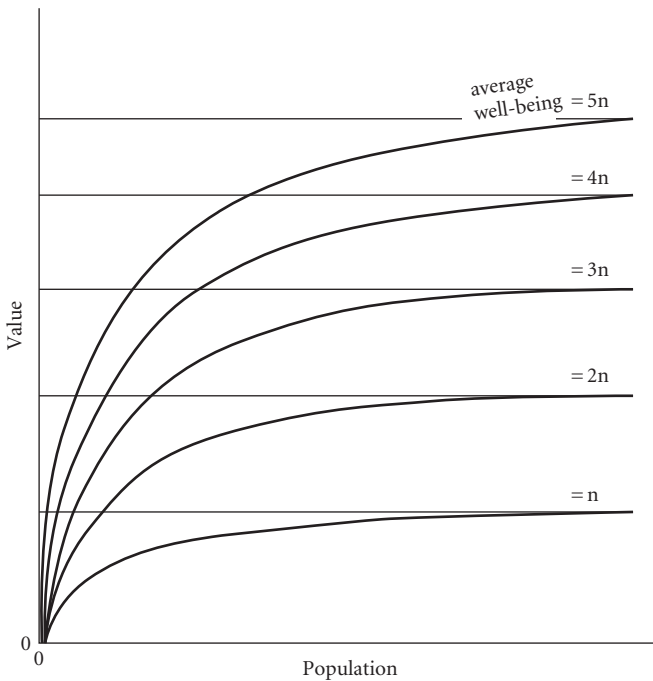


FIGURE 2.4 *Principle V2*

To approach these objections let us consider two worlds, which we can call W and $W+$. World W contains a certain number of people at a certain average well-being, while $W+$ contains the same people at the same average well-being, plus some extra people at a lower average well-being. Which of these two worlds is the better one? Since W has a higher average well-being the average principle will always prefer W , and since $W+$ has a greater total well-being the total principle will always prefer $W+$. But V_1 and V_2 will sometimes prefer W and sometimes prefer $W+$. If the population in W is very small, and the average well-being in $W+$ is not all that much lower than in W , V_1 and V_2 will probably prefer the larger world $W+$. But if the population in W is very large, and the average in $W+$ is quite a lot lower than in W , they will probably prefer the smaller world W .

Now if V_1 and V_2 sometimes prefer W to $W+$, they will sometimes forbid us to move from W to $W+$ by adding the extra people in $W+$. The first objection claims that this is implausible. World $W+$, it says, contains everything that W contains and differs from it only by a “mere addition.” And how can a mere addition make a world worse than it was before?

This mere-addition objection is often thought to have devastating force against the average principle,⁶ but there are two considerations that give it less force against V_1 and V_2 . The first is that V_1 and V_2 do not forbid nearly as many mere additions as the average principle does and, in particular, do not forbid the mere additions at low levels of population that it would be most implausible to forbid. This is especially true of V_2 . Since V_2 makes the value (and disvalue) of changes in the average well-being diminish to zero as the average well-being gets higher, it forbids many fewer mere additions at high averages than V_1 does, and as far as I can see, it hardly forbids any mere additions at high averages at all. The only mere additions that V_2 clearly forbids are those that start from a fairly large population at a fairly low average well-being per person—and these are the mere additions that I think it is the least implausible of all to forbid.

The second consideration only applies if we interpret “well-being” not as happiness but as involving the achievement of some form of human perfection, as Aquinas, Leibniz, and Kant all interpreted it, and as I would want to interpret it as well. When we make judgments involving perfectionist (as opposed to utilitarian) values, we often do object to (at least some) mere additions. Consider, for instance, the judgments we make about careers. Many of us think Muhammad Ali’s boxing career, to take a well-known example, would have been better without those last fights against Larry Holmes and Trevor Berbick. This is not because we think Ali’s performances against Holmes and Berbick were by some objective standard bad; we know that, for many other boxers, to do as well as Ali did against these fighters would have marked the pinnacle of their careers. It is rather because

⁶ E.g., by R. I. Sikora, “Is It Wrong to Prevent the Existence of Future Generations?” in R. I. Sikora and Brian Barry, eds., *Obligations to Future Generations* (Philadelphia: Temple University Press, 1978), p. 116.

we think Ali's performances were so much worse than the performances he produced in his prime that it was bad *for him* to produce them. The Holmes and Berbeck fights were mere additions to Ali's boxing career, yet many of us think his career would have been better without them. A similar attitude is present in the judgments many of us make about collections. Consider a collection of 100 exceptionally fine paintings, and then consider the collection that results when it is expanded by the purchase of twenty-five utterly mediocre paintings. If we think the second collection is worse than the first (as I am inclined to myself), then we should not automatically object to a theory that says that, given a world containing 1 billion people leading active, challenging, and autonomous lives, we make that world worse if we add to it 250 million extra people leading mindless, passive, and conditioned lives.

I would not want to say that these two considerations provide a complete answer to the mere addition objection, even for the most favored principle—namely, a perfectionist version of V₂. But I do think that together they provide a fairly good answer; and when we remember that the cost of providing a complete answer is accepting the repugnant conclusion, I think they provide an answer that is more than good enough.

If V₁ and V₂ sometimes prefer W to W₊, they will also, if taken on their own, sometimes require us to move from W₊ to W by killing the extra people in W₊ or by allowing them to die if they are in danger from which we could save them. The second objection claims that these consequences, which concern population decrease rather than population increase, are simply unacceptable.

This second objection is also often thought to have devastating force against the average principle, and it is certainly not an objection that we can answer by showing that V₁ and V₂ require a little less killing than the average principle or by pointing to some differences between perfectionist and utilitarian values. If I still do not think the objection has much force against V₁ and V₂, it is because I think consequences of the kind it points to are not peculiar to V₁ and V₂ but will follow from any consequentialist principle if we try to take it on its own. Consider, for instance, the total principle. Although this principle gives much more weight to population increases at high levels than V₁ and V₂, it still has notoriously unattractive consequences about killing and allowing to die. As its critics often point out, the total principle taken on its own would require us to kill innocent persons or allow them to die whenever this enabled us to replace them with slightly better off persons, and it would require killing and allowing to die in many other unacceptable circumstances as well. Some adherents of the total principle have tried to avoid these unattractive consequences by adopting a supplementary principle forbidding us to interfere with the autonomy of others,⁷ but while this autonomy

⁷ See, e.g., Jonathan Glover, *Causing Death and Saving Lives* (Harmondsworth: Penguin Books, 1977), pp. 71–72; and Peter Singer, *Practical Ethics* (Cambridge: Cambridge University Press, 1979), pp. 83–84.

principle may avoid some of the consequences about killing it does not avoid those about allowing to die. I am not sure exactly what supplementary principles are required if a consequentialist principle is not to have unacceptable consequences about killing and allowing to die, but I am sure that, whatever they are, they are required just as much by the total principle as they are by V_1 and V_2 . It is foolish to think that the consequentialist principles we use to assess the values of different populations could ever be the only principles in an acceptable moral theory. They have to be accompanied by supplementary principles setting constraints that we must not violate while pursuing our population goals and that we must not violate in particular by taking the lives of existing people. If we are to assess population principles *as population principles*, then we must assess them in circumstances where these constraints do not apply—that is, in circumstances where only increases and not decreases in the human population are in question. And when we do assess them in these circumstances, I think we find that V_1 and V_2 are the most attractive population principles that can be devised.

Postscript

The idea of representing this essay's main view on the specific graphs it uses was suggested to me by John Leslie, who also pointed me to the quotations from Aquinas, Leibniz, and Kant. What I then called the "variable value view" I would now call a "diminishing marginal value view," since it makes the value of additional people get smaller the more people there are. The essay considers only populations with positive average well-being (for example, positive happiness), and that is an omission—what about populations that suffer pain? Here it would not be plausible to say that, once there are billions of people suffering intense pain, it does not matter much if there are additional people in pain; the relation of value to size in such populations must be linear, with straight lines below the origin. This makes for a further asymmetry between pleasure and pain alongside the one discussed in chapter 7 ("Asymmetries in Value"); it also implies that a world with some immense number of people enjoying lives of great positive value can be on balance evil if sprinkled among them is a much smaller number of people suffering pain. This follows because, while the positive value in the majority's lives cannot exceed some limit, the negative value in the suffering lives has no limit. But this does not strike me as unacceptable: it is just another reflection of the way pain is a greater evil than pleasure is a good.

The Well-Rounded Life

This chapter discusses a moral problem, but it is not the most important moral problem in the world. It is not nuclear deterrence, abortion, or the relief of starvation. Still, it is a problem which I hope readers find engaging, and on which I hope they have views.

The problem arises within the moral view called perfectionism, so I must first explain what perfectionism is.

I. Aristotelian Perfectionism

Perfectionism holds that there are certain intrinsically valuable states of human beings, and that each of us should maximize the achievement of these states both in herself and in others. Like utilitarianism, then, it is a species of consequentialism. It thinks the morally right act is the one which produces the most good. But perfectionism differs from utilitarianism about what this good is. Whereas utilitarians identify the good with happiness, perfectionists find it in such states as knowledge, artistic creation (or contemplation), the carrying out of complex and difficult projects, and friendship. They want us to lead the best life, and to make its conditions available to others. But what determines the best life is not pleasure or satisfaction; it is the achievement of certain objective goods. Perfectionism is the morality of Plato, Aristotle, Aquinas, G. W. Leibniz, Hegel, Marx, Nietzsche, F. H. Bradley, and (in part) G. E. Moore and Hastings Rashdall. So it has been prominent in our tradition. But, for several bad reasons, it has been ignored by philosophers since about 1920. It is time for a reconsideration.

For their help some years ago, my thanks to R. M. Hare, Gavin Lawrence, Michael Lockwood, Amartya Sen, and Charles Taylor. For more recent suggestions, the same to G. A. Cohen, David Copp, Dennis McKerlie, Christine McKinnon, Derek Parfit, Stephen Perry, Donald Regan, and Terry Teskey.

Perfectionism might seem less important, if the activities deemed best by its lights were always most satisfying or brought the most happiness. In various forms, this has been held by Aristotle, J. S. Mill, and John Rawls. But the view is surely naive. It is a commonplace that knowledge can be painful, and serious achievement a strain. An artist's life, balancing joy in successful creation with the frequent anguish of failure, can be less contented overall than the life of an idle clubman. And it is clearly less contented than life in *Brave New World*. So it matters that perfectionism finds it better. And, even when good activities are most satisfying, a question of theory remains. Are the activities good because they are perfect, or because they are pleasant? Here perfectionism says, as Bradley did, that, even if life without pleasure is inconceivable, "what we hold to against every possible modification of Hedonism, is that the standard and test is in higher and lower function, not in more or less pleasure."¹

The problem I want to discuss can arise in any perfectionism, but it is most usefully discussed with a specific version in view. So let me briefly describe what I consider the most plausible perfectionism, a theory called *Aristotelian perfectionism*.

Like Aristotle's, this theory starts from the idea that, at the deepest level, the intrinsic goods develop properties fundamental to human nature. A little vaguely, they develop properties essential to us qua human organisms. This idea generates three major goods, of which the first is *physical perfection*. This perfection develops the bodily properties we share with other animals, and is present to a low degree in ordinary good health, and more completely in vigorous athletic activity, especially at the highest, perhaps Olympic levels. So the first Aristotelian claim is that a life is better and more completely human, if it is free of disease or injury and marked by some demanding physical feats. The remaining goods—the more important ones, it will emerge—are exercises of rationality, both theoretical and practical. On the theoretical side, the side of our beliefs, we achieve more *theoretical perfection* the more knowledge we have. The more we understand the world, and ourselves, and our place in the world, the better and more choice-worthy our lives. Of course, not all knowledge has equal value. Knowing the costars in some 1930s movie is not as important as knowing a fundamental law of the universe, or understanding the workings of a friend's personality. We need a test for the best knowledge, and I suggest it is the most organized or systematic knowledge, with general principles unifying and explaining derived particulars. This is most clearly present when we grasp a whole scientific theory from first principles down to particular explanations. But it is also present in interpersonal understanding and even the craft knowledge of skilled artisans. In any case, the second Aristotelian claim is that a life is better the more it is informed and aware, and the more it has unified its

¹ "Mr. Sidgwick's Hedonism," in *Collected Essays*, vol. 1 (New York: Oxford University Press, 1935), p. 97.

knowledge in one structure. The last good is *practical perfection*, or the exercise of rationality in action. Parallel to knowledge is the successful achievement of one's goals given a justified belief that this would happen, or, more briefly, nonlucky achievement. This is the basic practical value, but, again, some achievements are more valuable and exhibit more rationality than others. Just as it is better to know a scientific law than to know about 1930s movies, so it is better to achieve a fundamental reform of one's society than to light a cigarette successfully. The question of how goals are ranked is difficult, but I would say the best achievements involve the same organizing structure in a person's intentions as is present in the best knowledge. On this view, there is practical perfection in any life which is organized around a single end, or in which large parts have one end, so many activities serve a long-term goal; in activities that are complex, challenging, and difficult, especially if they require precise, and precisely ordered decisions; in political leadership and cooperation with others, which spread one's concerns beyond the self; and in friendship, especially when it is expressed in a nuanced emotional responsiveness. These are all practical excellences or exercises of rationality in action, and they make a life better by their presence.

Much more could be said about each of these goods, but I want to discuss something different. In our lives we often face choices between the goods. On a particular afternoon we may be able to read a book on European history or campaign for political office. Here we can increase our theoretical or practical perfection but not both, or not both equally. More generally, we must choose between a life of predominantly theoretical achievement—an intellectual life, if you like—a life devoted to action, and a life that tries to balance the two. In these cases, we have before us two ideals: knowledge and action, or the thinker and the doer. Should we prefer one ideal to the other, and, if so, which one? Or should we try to combine the two? And the point is quite general. Whenever there are distinct or plural perfections, we will face choices between them, both at particular moments and for our lives as a whole. After defining its individual goods, a perfectionist morality must help us make these choices. It must say how our overall perfection arises out of the individual perfections we achieve, or, more technically, how we can compare perfections when they conflict.

So far as I can see, the structure of perfectionism sets no limits here, nor does an Aristotelian view of human nature. Together they yield a list of perfections, but say nothing about how they are weighed. Nothing in the bowels of perfectionism requires one kind of comparison over others, and the strongest perfectionism can use whatever comparative approach seems most attractive.

II Lexicographic and Constant Comparisons

For many perfectionists this approach is very simple. They say there is one perfection, often theoretical perfection, which we should always seek ahead of others. If there are

different duties to pursue different goods, they give one of them absolute or *lexicographic* priority, so its demands always have moral precedence. Aristotle takes this line when he ranks contemplation ahead of any practical perfection. He wants us to “strain every nerve” to develop the best part of us, namely the intellectual part, implying that we should prefer any intellectual activity to the exercise of practical virtue.² Aquinas has a similar view about the relation between bodily development and rational perfection in general. Unlike Aristotle, Aquinas thinks bodily development has some value,³ but this value is “ordained to” spiritual goods and must always yield before them.⁴ Both writers make one perfection infinitely better than the rest, so even a tiny amount of it outweighs a great achievement of the others. Only when we can no longer promote the higher excellence should we seek a lesser good.⁵

Common though it is, this lexicographic approach is too extreme to be plausible. Each Aristotelian perfection has independent attraction, and none deserves to be placed so far above or below the others. Physical perfection seems to me less important than the two rational excellences. I have no special argument for this, but the development of our bodily powers seems to have less value than the exercise of our distinctive rationality. So it should count for less. But to make any increase in rationality, no matter how trivial, outweigh the greatest athletic achievements is to go too far. A plausible morality must give each of the Aristotelian goods some serious weight.

The simplest such morality employs *constant tradeoffs* between the different perfections. It assigns a fixed, finite weight to each excellence, and this weight is the same for all persons in all situations. If it makes 2 units of theory equal 3 of practice, it does so uniformly. Not only must I weigh theory and practice in this proportion, so must you and everyone else.⁶ Because there is no pre-existing scale on which perfections are measured, this view cannot in the strict sense make one higher. (If the units for theory were halved, the above proportion of 2:3 would

² *Nicomachean Ethics*, 1177b26–1178a7; also *Eudemian Ethics*, 1249b6–25.

³ According to Aquinas, the perfection of our bodily parts is a necessary component in the blessed state that will follow the resurrection of the body (*Summa Theologica*, I–II, q. 3, a. 3; q. 4, a. 5; also Supp., qq. 82–84), and also characterized the (slightly less perfect) state that preceded the Fall. Arguing against Origen that humans would indeed have reproduced by sexual intercourse had they remained longer in a state of innocence, Aquinas adds that this intercourse would have been vastly more pleasurable than anything we experience now, because of “the greater purity of nature and the greater sensibility of the body” (*ibid.*, I, q. 98, a. 2).

⁴ *Summa Theologica*, ii–ii, q. 152, a. 2, 3.

⁵ A theory with this view need not require us always to seek the highest excellence in ourselves. If it is agent neutral, as suggested above, it can tell us to sacrifice some of our own achievement of this excellence to promote it in others. This seems to be the view of Aquinas. While agreeing with Aristotle that contemplation is in itself better than any practical good (*Summa Theologica*, II–II, q. 182, a. 1), Aquinas thinks it meritorious to abandon one’s own contemplation to promote it in others (*ibid.*, II–II, q. 182, a. 2, q. 184, a. 7; also III, q. 40, a. 1, ad 1). Since “it is better to illumine than merely to shine” (*ibid.*, II–II, q. 188, a. 6), the most praiseworthy lives are devoted, as Aquinas’s was, to teaching as well as to study.

⁶ Nothing in these remarks implies that comparisons can realistically be made with such precision, so that they are 2:3 rather than, say, 21:30. Nor does it imply that the individual excellences can be cardinally measured, as the mere possibility of such precision requires. The idea is just that we can deepen our understanding of comparative judgment by asking what form precise tradeoffs would have if they were possible.

become 4:3.) But it can do this informally. If our comparative principles imply that most people should spend more time pursuing one good than another, there is a clear sense in which they make it more important. This is just what we should want for rational versus physical excellence.

Constant tradeoffs still permit Aristotle's view that theory is better than practice, now read as the nonlexicographic claim that it deserves more (finite) weight. But, even in this weaker form, the view is implausible. Theoretical and practical excellence both realize rationality, and do so in structurally similar ways. Both are products of evolution, and nothing in their character or origins makes one more desirable than the other. Why should rationality in conduct, in how we change the world, count less than rationality in how we form our beliefs? Why should a structure of ends have less value than a similar structure of convictions? Aristotle's arguments for preferring theory are unimpressive. His central argument—that theoretical excellence realizes a separable and divine element in our nature⁷—is of merely historical interest, as is the claim that it concerns the best objects.⁸ Remove God and we have no reason to think what theory knows better than what action affects. Nor are his other arguments more persuasive. That contemplation offers pure and continuous pleasure is irrelevant to perfectionism; that it is more self-sufficient is dubious; and that it alone is loved for itself is false.⁹ Skill in games is valued not just as a means to winning—many of us care little for that—but because in itself it exercises rational capacities. Just like contemplation, it can be the object of intrinsic appreciation. Without better arguments than Aristotle's, perfectionism should give the rational excellences roughly equal weight, so a world of uniformly good lives devotes roughly equal time to each.

III. Balancing

Our initial suggestion, then, is to have constant tradeoffs between the excellences, with physical perfection ranked lower and the rational goods roughly equal. This is not positively unattractive, but I think many of us are drawn to something different. We think the best lives contain a certain balance between the perfections, and do not concentrate too much on one. To talk only of Leonardo would pitch the ideal too high, for what I intend is possible in all lives. We all can spread our activities between different goods, aiming at a well-rounded achievement rather than any narrow specialization. Even if our specific achievements are not great, their proportion can mirror that of Renaissance lives, and for many of us this proportion is, other things being equal, a good. At however modest a level, it gives our lives an intrinsically desirable shape.

⁷ *Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984), 1177b26–1178a8.

⁸ *Nicomachean Ethics*, 1177a21.

⁹ For these three arguments, see *Nicomachean Ethics*, 1177a24–b25; and, for a reprise, Aquinas, *Summa Theologica*, II–II, q. 182, a. 1.

This *balancing* view, however, excludes constant tradeoffs. Imagine that one person has devoted most of her life to politics, while another has worked only at scholarship. Well-roundedness implies that the first will improve her life most by acquiring some knowledge, while the second should become more active. They should prefer what they have neglected, and what they have neglected is different. At the heart of balancing is the idea that a perfection's relative value depends on the relative amount of it one has achieved in the past. Going beyond equal weights, it says that, if one excellence has been achieved more than another, the second is more important. The clearest representation of this is on an indifference graph (Fig. 3.1). Here each curve links points representing combinations of the two goods that make for the same overall perfection, with curves farther from the origin better than ones close in. For simplicity, imagine the axes calibrated so that one unit of each perfection is what an average person can achieve in a fixed time, say one hour:

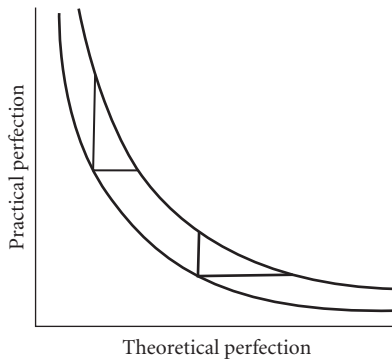


FIGURE 3.1 *Balancing Curves*

In the upper left of the graph, where the politician has achieved more practice than theory, the same overall improvement results from a large increase in action or a small increase in knowledge. So, for her, knowledge is more valuable. In the lower right, by contrast, the scholar makes an identical gain by improving theory greatly or practice a little. Given different past achievements, there are different values now.

Indifference curves are borrowed from economics, where they are used to represent consumer preferences among different commodities. But there is an important difference between the economists' use and our own. On a natural view, the economists' curves reflect the fact that the absolute utilities of commodities diminish as their absolute levels increase. Thus, the reason we prefer an apple to an orange when we have twice as many oranges as apples is that we care less for either fruit the more of it we have. This cannot be the point in perfectionism. Imagine that one person has 10 units each of theoretical and practical perfection, while another has 20. If the value of balanced lives rested on diminishing *absolute* values, the second person's life would be less than twice as good as the first's, which we do not want to say. Our aim is to appreciate Leonardo, not to minimize his feats. For this reason, acceptable

balancing principles must look only at a person's *relative* achievements, so ratios between the goods, and not their absolute levels, are what matter.

Given the intuitive appeal of balancing, we should not be surprised that it appears frequently in the perfectionist tradition. Several writers say that an ideal life contains a "harmonious" achievement of different goods. Sir William Hamilton defines perfection as "the full and harmonious development of all our faculties";¹⁰ Wilhelm von Humboldt calls a human's true end "the highest and most harmonious development of his powers to a complete and consistent whole";¹¹ and Rashdall describes a life where "many distinguishable elements are harmonized and combined."¹² At times the point of these remarks is causal. Humboldt says a person can "increase and diversify the powers with which he works, by harmoniously combining them, instead of looking for a mere variety of objects for their separate exercise."¹³ This is sound advice on any view of comparison. If some activities, or ways of organizing activities, enable us to achieve two excellences at once, any perfectionism should find them good. But the talk of harmony goes beyond this, and suggests that some proportion between goods is violated if we concentrate too much on one. The same view is present in the polemics of Marx and Nietzsche against specialization and the division of labor. Marx famously wants to be able "to do one thing today and another tomorrow, to hunt in the morning, fish in the afternoon, rear cattle in the evening, criticize after dinner, just as I have a mind, without becoming hunter, fisherman, cowherd, or critic";¹⁴ while Nietzsche heaps scorn on the "nook-dwellers" and "fragments of humanity" he finds among European intellectuals.¹⁵ Against "modern ideas" that would "banish everybody into a corner and 'speciality,'" he insists that man's greatness lies in his "range and multiplicity, in his wholeness in manifoldness."¹⁶

The exact impact of a balancing view depends on facts about the lives we have available—facts we discuss below (sec. IV). But it also depends on the goods' weighting. Like constant comparison, balancing can make one perfection higher if it tells us to spend more time pursuing it. (This will be reflected in curves that tilt toward one axis, rather than lie symmetrically as in Figure 3.1.) We should take this line with rational and physical perfection. Then our

¹⁰ *The Metaphysics of Sir William Hamilton*, ed. F. Bowen (Boston: John Allyn, 1872), p. 1.

¹¹ *The Limits of State Action*, trans. J. W. Burrow (New York: Cambridge University Press, 1969), p. 16.

¹² *The Theory of Good and Evil*, vol. 1 (Oxford: Oxford University Press, 1907), pp. 220–21. Elsewhere Rashdall says that "no amount of one good can compensate for the absence or deficiency of the other" (vol. 2, p. 39), an even clearer balancing claim.

¹³ *The Limits of State Action*, p. 16.

¹⁴ Marx and Friedrich Engels, *The German Ideology*, in D. McLellan, ed., *Karl Marx: Selected Writings* (New York: Oxford University Press, 1977), p. 169; see also Marx, *Grundrisse*, trans. M. Nicolaus (Harmondsworth: Penguin, 1973), pp. 325, 487–8.

¹⁵ *Thus Spake Zarathustra*, Second Part, "On Redemption," in W. Kaufmann, ed., *The Portable Nietzsche* (Harmondsworth: Penguin, 1976), p. 250. Also *Beyond Good and Evil*, trans. W. Kaufmann (New York: Vintage, 1966), sec. 204, 205, 212, 257; *On the Genealogy of Morals*, trans. W. Kaufmann and R. J. Hollingdale (New York: Vintage, 1969), Third Essay, sec. 23; *Twilight of the Idols*, "What the Germans Lack," sec. 3, in *The Portable Nietzsche*, p. 508; *The Antichrist*, sec. 57, *ibid.*, p. 647; and *The Will to Power*, trans. W. Kaufmann (New York: Vintage, 1968), secs. 390, 881.

¹⁶ *Beyond Good and Evil*, sec. 212.

theory will say that, although it is important for a life to contain some bodily goods, there should usually be less time spent on them than on rationality.

Balancing has several indirect consequences, one of which is to reduce the importance of an excited chase after goods. On any view, the best lives strive vigorously after excellence. But, with balancing, this does not guarantee a close approach to overall perfection. There are many lives of all-out activity that fail, through lack of balance, to equal the value of lazier lives whose proportion is right. Energy matters, but, given balancing, its level can be less important than the direction in which it is expended. At the same time, balancing increases the importance of self-knowledge. With constant tradeoffs, a person can compare without knowledge of her past. Since the relative weights of the perfections are always the same, she can choose correctly without knowing how she chose before. Balancing, however, presupposes an awareness of one's history. As a realization of theory, this awareness is intrinsically good, and, by contributing to a life that is unified across time, it also advances practice. Here we see how, given balancing, self-knowledge is also a condition for correct comparative judgment.

Alongside its consequences for the general shape of our lives, balancing affects our choice of specific activities. As Humboldt points out, one way to achieve "harmonious development" is to choose projects that promote several excellences at once. These projects are recommended by any perfectionism, but even more by one with balancing. Scientific research passes this test, for it combines the goods of knowledge and action. A scientific researcher sets herself a goal—discovering a new law, or uncovering the motion of continents—and uses skill and ingenuity to achieve it. Her activity applies and creates valuable knowledge, but it also involves subtle strategic thinking. Something similar can happen in athletics. Team sports can require players to grasp and solve strategic problems at the same time as they exercise their bodies. Here the mix of rational with physical goods is what balancing finds attractive, and what suits these sports to a well-rounded life.

Given this last point, a life could count as well-rounded, in the sense defined up to now, even though it concentrated on just one activity. A life devoted entirely to scientific research could be balanced if its inquiries developed practice as much as theory, and a narrowly hockey-playing life could be rounded if the sport's intellectual demands match those on the body. This may seem counterintuitive, and false to the original ideal of well-roundedness. It may seem far from honor to Leonardo to think his virtues attainable in a life with just one realm of achievement. The result is counterintuitive, but we can avoid it by extending the balancing view from different perfections to different realizations of one perfection. Then it will, other things being equal, be better to have knowledge of European history, astrophysics, and one's friends' characters than to concentrate understanding in a single area, and better to carry out complex plans of different types than to have just one kind of practical achievement. This extension of balancing is natural, and

I will assume it in what follows. I will assume that we are to seek variety, not just among different excellences, but among different aspects of one excellence; so a fully rounded life is developed, knowing, and active in several areas. Unfortunately, the extension complicates our view, and makes it harder to represent graphically. So I will often continue to speak as if we are to seek balance primarily among our three most general perfections.

IV. Dilettantism and Concentration

Balancing says that, other things being equal, we should prefer the excellences we have achieved less in the past. But, to apply this advice, we must know when other things are equal, or, more precisely, what other effects a search for balance will have. Does diversity of action undermine the specific goods, so a well-rounded life is one of uniform mediocrity? If so, balance, while attractive in theory, is not practically recommended. Or does diversity enhance the perfections, enriching them from different sources? Then balance is doubly valuable.

To sharpen these questions, let us return to our indifference graph and add an *achievement line*. The points on this line represent the greatest combinations of perfection a person can achieve given her natural abilities, her resources, the time in her life, and whatever else affects her excellence. The area under and including the line represents the total lives available to her, and, in comparing perfections, she in effect chooses a point in this area. If her achievement line is straight and symmetric to the axes (Fig. 3.2), it touches the outermost indifference curve at a single central point, which we can call its *ideal-life point*. Given this line, she leads her best life by giving equal time to both goods.

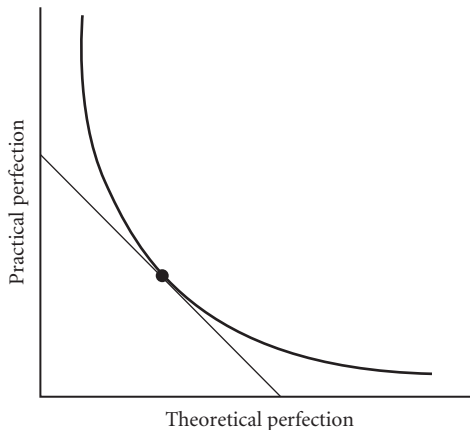


FIGURE 3.2 *Straight Achievement Line*

Real-world achievement lines, however, are hardly straight, and their shape can affect the impact of balancing. Imagine, first, a line that is convex to the origin, so its shape mirrors that of the indifference curves [line (i) in Fig. 3.3]:

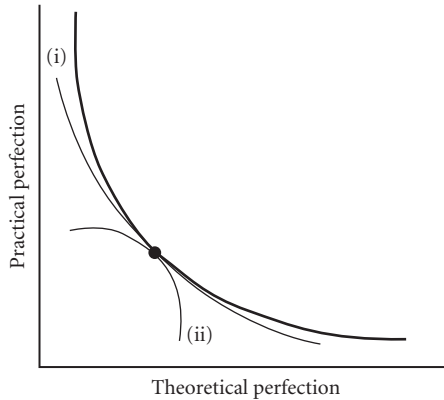


FIGURE 3.3 *Curved Achievement Lines*

With this line, balance is not such a virtue, nor specialization such a crime. Even very concentrated lives come close to the outermost indifference curve, and, with sufficient convexity, could touch it. By contrast, a line that is concave to the origin [line (ii) in Fig. 3.3] makes balance more important. Concentrated lives fall much shorter of the ideal, and are even more seriously wrong.

Achievement lines belong to individuals, and it would be foolish to think that one shape fits all. Nevertheless, there may be general factors that affect the shape of all lines and determine the general import of balancing. What could they be?

There seem, unfortunately, to be two opposing factors, one supporting convexity and one its opposite. In favor of the convex line, and against balance, is what we can call *dilettante's disadvantage*. People often find that the more time they invest in an activity, the more they gain from further attention to it. A professional historian can learn more from an hour with a history book than can an athlete, who lacks the background to read it critically. At the same time, the athlete can achieve more physical perfection in an hour's exercise. He can run farther, or lift greater weights. In both cases, the past devotion pays dividends now. It is often said that people who spread themselves across many activities will not achieve much in any. If this is correct, a well-rounded life will be low in all perfections, and, to reflect this, most achievement lines will be convex.

On the other side are the *costs of concentration*. People who devote themselves to one activity can easily become fatigued or even bored. They can miss the invigorating effect of variety, as well as the chance to enrich themselves in one area with experiences gained doing something different. Their very narrowness can produce stultification. What is more, some perfections seem subject to a law of diminishing returns, which makes it easier to move from low to middling achievements than

from there to the highest heights. In ballet, for instance, it takes a certain effort to move from a beginner's level to that of dancing in the corps of a university production. But it takes much more effort to move from there to a professional production. At the highest, perhaps Bolshoi, levels, dancers practice for hours to improve their movements a little. Given their tremendous costs, might they not do better for themselves if they sought some variety? The worry here is not that the dancers are wrong to practice as much as they do. Given the finer appreciations they permit the thousands of people in their audiences, they are probably, on balance, right. But, if the cost of these appreciations is such narrow concentration by the dancers, may it not leave their own lives impoverished?

These opposing factors are both plausible, and readers may disagree about their relative weights. Some may think the costs of concentration more important, and continue to value well-roundedness even when all is considered. Others may emphasize the disadvantage. But there is a third possibility. We may be able to acknowledge both factors and unite them in one view, if we look more closely at where their force is greatest. Although they pull in different directions, the factors seem to pull most strongly in different places, and may therefore not be directly contradictory.

Dilettante's disadvantage, to start, seems strongest at low levels of perfection and weaker high up. Many arenas of perfection are such that large investments of time are needed before any value results. In a cognitive domain, there are basic concepts and principles to grasp; in a practice, there are fundamental skills, and without them a person's activity counts little. Unless he does enough to acquire the fundamentals, his achievement is minimal. (Consider learning a language: without a basic grasp of the grammar, individual bits of vocabulary mean little.) Once these fundamentals are present, however, new possibilities open. A person can use the basic concepts and skills to acquire others, and, for a time, his progress is rapid. Because of this, diversity is most damaging when it prevents a person from acquiring any fundamentals. But it is less so farther along. Someone with the elements of several perfections can make real gains in them all, and need not be distracted by variety. Where a beginner must concentrate, the advanced practitioner can alternate among activities, returning to each with his mastery secure.

The costs of concentration, by contrast, seem greatest at high levels. Someone who devotes an hour a day to history is in little danger of fatigue, and can easily add an extra hour. Doing so may even increase his scholarly commitment. He may have more energy given two hours a day than if he is only able to dabble. After eight hours, however, things become different. Now boredom and fatigue are real factors, and it is questionable how much someone studying that hard can add in an extra hour per day. It might even improve his history to relax with something different. What is more, if diminishing returns affect some perfections, they should do so primarily at high levels. And this does seem plausible. In many domains, the rapid progress that follows the acquisition of the fundamentals

eventually slows. As a person approaches the frontiers of his discipline, further advances become more difficult, requiring more effort and application. The return of perfection to time becomes less, not more.¹⁷ Finally, there are interaction effects. Although diversity is distracting at low levels, it can be fruitful farther on. Major new advances often occur when insights from one area are applied to another, and those who seek such advances will do well to broaden their experience.

These arguments suggest a unified picture: dilettante's disadvantage is strongest at low levels, and the costs of concentration greatest higher up. The bad effects of variety are most significant when one's achievements are low, and the costs of specializing worst when they are high. And this, in turn, suggests that many people's achievement lines may be M-shaped, as in Figure 3.4:

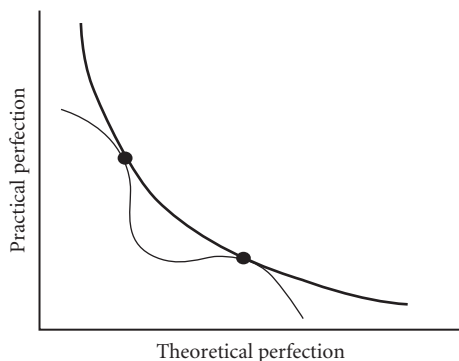


FIGURE 3.4 *M-Shaped Achievement Lines*

In the upper left and lower right of Figure 3.4, where the achievement of one good is quite high, the costs of concentration work strongly on that good to make the line concave. In the center, however, where both perfections are low, the disadvantage dominates and the line is convex. If the M shape characterizes a person's options, she has two better lives available, each concentrating to some extent on one good. She must be neither a pure specialist nor a pure all-rounder, but give a moderate preponderance to one perfection. To give more than that preponderance is to fall afoul of the costs, which can be serious. But it is also bad to aim at too much balance. Someone who does that will slide into the central trough of her achievement line, and onto a lower curve.

¹⁷ It is this possibility—different returns of perfection at different levels, as well as to people with different talents—that forces us to equate one unit of perfection with what an *average* person achieves in an hour (sec. III). With cardinal measures of the individual excellences, we can compare goods in one area, e.g., Leonardo's paintings with a beginner's. But we need to normalize our measures, and, for this, look to an average person's achievements in a fixed time.

In the configuration of Figure 3.4, the M shape has most importance, and the upshot would not be very different with constant tradeoffs. But balancing still plays a role. For one thing, it makes the ideal life slightly more balanced, with its point on the inside slope of a hump rather than right on top. So a person's goal is slightly more proportion. Balancing also affects the badness of less than ideal lives. Because the indifference curves dip into the central trough, a dilettantish life is, while not best, less bad than if the curves were straight. And an overly specialized life is worse. It falls much shorter of the outermost curve, and is more seriously lacking. Even with M lines, balancing affects deliberation, pushing us toward proportion and away from monomania.

When a person has two ideal life points, as in Figure 3.4, she can choose between two equally good lives, and may do so as she pleases. With precise indifference curves, this equality is unlikely, since most people have somewhat skewed abilities. Their talents incline them somewhat more to one perfection. But, given the many curves in a realistic perfectionism—or, what is the same thing, given the merely partial comparability of its values (see note 6)—something like it may be common. Many people may have several good lives available, each concentrating moderately on one good, and be morally free to choose among them. Having made an initial decision, they must stick to it, for otherwise they will lapse into dilettantism. So past a point they must be firm in their commitment. But, at the start, their choice of lives is morally unconstrained.

If I am right about the (rough) shape of (many) achievement lines, they answer the most common objection against balancing, namely that it promotes mediocrity. Rashdall urges this objection against a “self-realization” theory that identifies good with “an equal, all-round development of one's whole nature”:

Up to a certain point the man who is a mere specialist will be a bad specialist, but that point is soon reached. Charles Darwin found that the cultivation of reasoning power and observation had extinguished his once keen imagination and sensibility. And yet who would wish—whether in the interests of the world or in the interests of what was best in Charles Darwin's own nature—that his work should have been spoiled in order that one of the three hours which was the maximum working day his health allowed should have been absorbed by politics or philanthropy? Who would decide that the origin of species should have been undiscovered, in order that the man who might have discovered it should retain the power of enjoying Wordsworth? This notion of an equal, all-round, “harmonious” development is thus a sheer impossibility, excluded by the very constitution of human nature, and incompatible with the welfare of society. And, in so far as some approximation to such an ideal of life is possible, it involves a very apotheosis of mediocrity, ineffectiveness, dilettantism.¹⁸

¹⁸ *The Theory of Good and Evil*, vol. 2, p. 64. How this argument is supposed to mesh with Rashdall's own remarks about “harmonizing” different elements (see above, note 12) I cannot say.

This entertaining passage rather exaggerates its history. Darwin “discovered” the theory of evolution some twenty years before publishing it in *The Origin of Species*, and continued scientific work for another twenty years after *The Origin of Species*. So a realistic picture of his life without an hour’s biology a day will have it lack not the discoveries that made him famous, but perhaps the works on *Insectivorous Plants* and *The Effects of Cross- and Self-Fertilization* which he published late in his career. And is it wrong to say that his life would have been more complete had he left the work of these books to others, and done more outside biology? I do not think so, and neither, it appears, did Darwin. In a memoir published by his son in 1929, Darwin is reported to have told one of his daughters that “if he had his life to live over again he would make it a rule to let no day go by without reading a few lines of poetry. Then he quietly added that he wished he had not ‘let his mind rot so.’”¹⁹

The Darwin example may be used to make a stronger claim than Rashdall’s. Some may say that if Darwin’s greatest talent was for science, he should have preferred even tiny achievements there to large ones elsewhere. He was above all a biologist, they may argue, and, in assessing his life, we should count his biological perfections above all others. The view advanced here is not the lexicographic one discussed earlier (sec. 11). It does not say that some activities are better for all humans, regardless of their talents or situation. But the view still supports specializing conclusions. It holds that, although different perfections are in themselves equally good, the value of a person’s life depends primarily on what he achieves in his single best one. If his talents incline him more to one good than to others, he should prefer any gain there to improvements in its rivals.

I have little to say about this *pure specializing* view. It directly counters the position I have been developing, and violates the intuitions I most want to capture. Although it is intelligible, I do not think it really survives the application to serious judgments. When we think seriously about lives, the compelling ideal is surely of well-roundedness. And we can retain this ideal if the most common objection to it, namely Rashdall’s objection, is answered by M lines. These lines make too much balance a bad thing; and views that value well-roundedness in the abstract can agree that, if appreciating Wordsworth had prevented Darwin’s main discoveries, it would have been wrong. At the same time, they can say that, of the broadly specialist lives available to Darwin, the best contained at least some other goods and were not solely devoted to biology.

I think the combination of balancing and M lines yields a perfectionism in harmony with many intuitions about specialization. On the one hand, it makes well-roundedness an ideal, and favors lives with the greatest balance consistent with serious achievement. On the other, it reflects our common suspicion of dilettantes.

¹⁹ Major Leonard Darwin, “Memories of Down House,” *The Nineteenth Century*, 106 (1929): 119–20, quoted in William Irvine, *Apes, Angels, and Victorians* (New York: Time, 1963), p. 147.

The combination may also help historically. When classical writers discuss theory and practice, they often compare, not knowledge and action as such, but the value of whole lives devoted to them. They do not ask whether understanding is better than political rule (for many the highest practical activity), but whether a philosophical or contemplative life is better than a political life.²⁰ This makes sense on the view we have described. It follows from balancing that we cannot compare two excellences directly. Since a unit of theory that would mean a great deal to a politician will add little to a scholar, we cannot assess it apart from the life in which it figures. The smallest unit we can compare is the complete human life. And, if this explains the focus on lives, the restriction to specialist lives may follow from M lines. If we want to compare the best lives commonly available, as most perfectionists do, the disadvantage tells us to ignore any that divide their time equally. We can consider contemplative and political lives, but not ones that aim at both goods. One perfectionist who discusses these lives is Plato, in his account of the philosopher-kings. But what Plato says is that, precisely because they have to combine such different abilities, these kings will be hard to find.²¹

A final point is this. M lines are plausible because dilettante's disadvantage is strongest at low levels and the costs of concentration greatest higher up. But, if this applies along one achievement line, it should also hold for a series of lines representing the options available to different people. Those with limited abilities or a short lifespan should find that dilettante's disadvantage works powerfully on their options, creating a deep central trough in their line. For those with more talent and time, however, this factor should be less important. As their achievement lines move farther from the origin, the troughs should become shallower, and the line should eventually approximate the simple concavity that would obtain given only costs of concentration (Fig. 3.5):

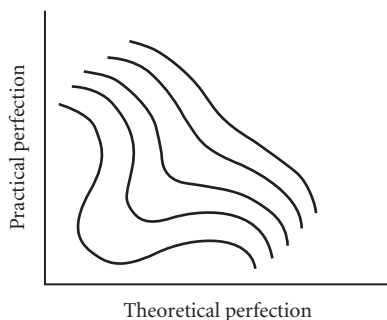


FIGURE 3.5 Series of M-Shaped Lines

²⁰ Aristotle, *Nicomachean Ethics*, 1095b14–1096a10, 1177a11–1179a33; *Eudemian Ethics*, 1215a26–1215b14, 1216a27–37. See also Aquinas, *Summa Theologica*, II–II, q. 182.

²¹ *Republic*, 503c.

This too matches intuition. For it implies that, although exceptional individuals like Marcus Aurelius and Leonardo can aim at well-rounded lives, those with more limited gifts must content themselves with being some kind of specialist.

V. Many-Person Balancing

So far we have applied balancing within single lives, and made the relative weight of a perfection depend on the relative degree to which a person herself has achieved it. But it is also possible to balance across lives. We can define an ideal proportion for a collective, and say a group does better when its members achieve a variety of goods than when they all specialize alike. Just as we value well-roundedness within lives, so we can value it across them, and prefer groups whose members exercise a diversity of talents.

For some perfectionists, this many-person balancing is a simple extension of the one-person approach. In his account of “harmonious development,” Humboldt says that what occurs in one person “by the union of the past and future with the present, is produced in society by the mutual co-operation of its members.” An individual can only achieve some excellences open to the species, but by entering into a “social union” with others she can overcome this limitation and “participate in the rich collective resources of all the others.”²² The intuitive idea here is that, if she belongs to a group whose members achieve all the goods, she can participate through them in a well-roundedness denied her on her own. But the idea again rests on balancing. A group can only offer this participation if some members’ achieving one perfection permits others to neglect it, and gives them a reason to seek other goods instead.

Far from extending one-person balancing, however, this view can conflict with it. Imagine that a person has achieved more theoretical than practical perfection, but belongs to a group whose other members have preferred practice. Balance in her life requires more action, while proportion in the group demands the opposite. The same conflict is possible for different groups. A person’s family may have achieved more theory than practice, but her country more practice than theory. So which should she prefer? Balance in different units makes different demands, and a theory that extends it faces difficult questions about where it takes precedence.

Because of these conflicts, a perfectionism with many-person balancing will be very complex, with a different overall goal for each collection where proportion is valued. This is not a decisive objection, but I think the strongest perfectionism should confine its balancing to single lives. Let me explain why.

We should agree, first, that the intuitive appeal of balancing is greatest in single lives. There an ideal of proportion strikes us immediately, whereas for larger groups it is less obvious. There is a reason for this. The degree to which balancing is plau-

²² *The Limits of State Action*, p. 17.

sible for a collection depends on the degree to which the collection forms a unity. The states making up a single human life are connected in especially intimate ways, and form a tighter unity than any collection across lives. This makes it especially natural to consider them together, and to seek proportion among them. To confirm this note that when we do balance in groups, it is only the most unified that we consider. We may praise a family if its members achieve different excellences, or admire a civilization like Greece for excelling in such diverse areas as statecraft, poetry, warfare, and philosophy. But, if we consider accidental collections like the group of people born on a Tuesday, or with names containing seven letters, the idea does not suggest itself. And it does so even less if a group has no internal connections. Imagine a universe containing two civilizations that live in galaxies light years apart and have no knowledge of each other. Surely it is not plausible to claim that, if one civilization is realizing theory more than practice, this gives the other grounds (which it cannot know) to concentrate instead on practice.²³

If this is right, decisions about balancing reflect views about where significant unities lie, or about what collections form significant units. Now, perfectionism has always made the complete life its central moral unit. Its first question has always been, “What is the best whole human life?”—not “What is the best state of a person at a time?” or “What is the best distribution of states across persons?” It would reflect this concern a little to balance everywhere, but to give the consequences in single lives the greatest weight. But it is surely simplest to stick to lives. If the single life is an important unit, we signal this most dramatically by confining balance to it.

And there is a wider connection. A striking claim of Rawls’s *A Theory of Justice* is that classical utilitarianism does not take seriously the separateness of persons,²⁴ or, looking at the same fact from the inside, the unity of individual lives. Different writers have used this claim differently, and to support different moralities;²⁵ but, for me, its core is this. The states morality deals with come divided into lives, and have more intimate relations within these lives than across them. This is an important metaphysical fact, and should be reflected in moral thought to the following extent. It should be impossible to apply a moral view conclusively without knowing how the states it deems important divide among persons. If all we know

²³ In a discussion that has influenced mine, Brian Barry suggests a balancing treatment of two social values that he calls equity and efficiency (*Political Argument* [London: Routledge & Kegan Paul, 1965], pp. 3–8; see also John Rawls, *A Theory of Justice* [Cambridge, Mass.: Harvard University Press, 1971], pp. 37–40). The proposal, however, is only plausible for something as integrated as one society. No one would claim that, if another nation is achieving more equity than efficiency, ours must seek more efficiency to redress the balance.

²⁴ *A Theory of Justice*, pp. 29, 187.

²⁵ In Rawls, separateness seems to support a methodology, namely the hypothetical contract, and only indirectly any egalitarianism chosen through it. But, more recently, Thomas Nagel and Samuel Scheffler use separateness to argue for equality directly (“Equality,” in *Mortal Questions* [New York: Cambridge University Press, 1979], pp. 106–27; and *The Rejection of Consequentialism* [New York: Oxford University

about an act is that it will produce such and such states, without knowing which go into which lives, we should not be able to judge the act decisively. This simple condition is not satisfied by classical utilitarianism, which can evaluate so long as it knows the total happiness an act produces. It does not take seriously the differences between persons, because it does not require us to know about these differences before we judge conduct.

Now, this reading of separateness also supports one-person balancing. We have already noted that this balancing prevents us from comparing excellences directly (sec. iv). An item of knowledge that would be good for a politician will add little to a scholar, and we cannot assess it apart from the life in which it figures. Perfectionist judgments presuppose comparison, and, with one-person balancing, this requires assigning goods to lives. The same is not true of many-person balancing, which can compare once it knows the total in a group. So, on this score, it is no better than classical utilitarianism. It would, again, meet Rawls's point a little to give one-person balance more weight than balance across lives, so it is only somewhat more important. But it is more dramatic to stick to lives. This restricts us to the balancing with the most intuitive appeal, and also acknowledges separateness in the strongest possible way.

The separateness of persons is important, but it does not require as radical a response as some writers suggest. To acknowledge the metaphysical facts, we need not abandon goal-based morality, as Nozick argues, or even summation across persons, as Rawls believes. We can retain a basically consequentialist structure so long as our goods are perfectionist, and, in weighing them, we consider only relative achievements within single lives.

Postscript

A longer version of this essay appeared as chapter 7 of my book *Perfectionism*.

Press, 1982], ch. 3). And Robert Nozick claims that separateness supports libertarianism, which notoriously ignores equality (*Anarchy, State, and Utopia* [New York: Basic Books, 1974], p. 33); John Taurek cites it in defending a view that is insensitive to numbers ("Should the Numbers Count?" *Philosophy and Public Affairs* 7 [1976–1977]: 293–316); and Derek Parfit thinks it may support rational egoism (*Reasons and Persons* [New York: Oxford University Press, 1984], pp. 139, 521). This array of arguments, all, it seems to me, equally plausible (or implausible), makes me doubt whether Rawls's claim has more force than that suggested for it in the text.

Monism, Pluralism, and Rational Regret

According to an increasingly widespread view, ethics concerns the evaluation not only of acts and intentions but also of attitudes and feelings. In different circumstances different attitudes can be appropriate, rational, or morally called for. As Aristotle held, the morally virtuous person not only acts and intends but also feels appropriately or well.

Much recent discussion of appropriate feelings has focused on a particular context: where you had a choice between acts that would produce different goods and chose one act, and one good, over the rest. In section I of this article I argue that the issues that arise in this context are but one instance of a more general issue about the division of appropriate attitudes. In section II I consider a claim often made about feelings after a choice: that regret for a forgone lesser good can be rational only given a pluralistic rather than a monistic theory of the good. Against this widely accepted view I argue that monism, too, allows for rational regret. In section III I consider the implications of this argument for our understanding of pluralism. Even if there are independent reasons to prefer pluralistic theories of the good, the argument supports moderate over extreme pluralism—that is, pluralism with a moderate rather than an immense number of generic goods.

I. Concentration Versus Proportionality

Imagine that you had a choice between two acts, one of which would produce a greater good and the other a lesser good. You chose the first act and produced the greater good. How is it rational, or appropriate, for you to feel about your act's

Earlier versions of this chapter were read in the spring of 1994 to the Conference on Incommensurability and Value in Colombiers-sur-Seulles, France, and the annual congress of the Canadian Philosophical Association in Calgary. For their helpful comments I am especially grateful to Ruth Chang, Roger Crisp, Ronald de Sousa, Tracy Isaacs, Shelly Kagan, Dennis McKerlie, Karen Pilkington, Donald Regan, and Michael Stocker.

result? The greater of the two goods now obtains, while the lesser does not. How should you feel about this fact?

A coherent, intelligible view says you should feel only and entirely pleased. You have produced a good, perhaps a great one, which is a reason for satisfaction. And although you have failed to produce another good, its being a lesser good means you have no reason to regret the effect of your choice. Having brought about the best outcome possible in your circumstances, you should feel only positively about that outcome.

This first view allows that it may be rational for you to feel some regret. Thus, it allows that you may and perhaps should wish that you had not had to choose between the goods but could instead have produced both simultaneously. The view allows, in other words, for what we can call regret about your choice situation. But it insists that, taking that situation as given, you should feel only pleased that you produced the greater good and not the lesser.

A different view agrees that, taking your situation as given, you should feel pleased at having produced the one, perhaps great, good. But it says it is also rational to regret not having produced the other, lesser good. Since the good you produced was the greater of the two possible for you, your pleasure should be more intense than your regret, so your predominant feeling is one of satisfaction. But it is still appropriate to feel some regret for the forgone lesser good. Like the first view, this second one can call for regret about your choice situation. But it holds in addition that you should feel some regret that, in your situation, you did not produce the lesser good.

These two views differ similarly about a case where you have somehow chosen the lesser of two goods. Here the first view says you should feel only regret for the forgone greater good and no pleasure at all in the lesser good. The second view says you should feel more regret than pleasure but still take some pleasure in the smaller good you did produce.

I will call these views the “concentration” and “proportionality” views. The concentration view says that, having chosen between two goods, you should direct all your feeling to the one that is greater. If you produced this greater good, you should feel only and entirely pleased; if you did not, you should feel only regret. The proportionality view says you should divide your feelings in proportion to the goods’ relative values: more for the greater good, but still some for the lesser good.¹ But in fact the issue between these views is more general than this initial introduction suggests.

¹ As I have characterized them, the concentration and proportionality views concern attitudes to the outcomes of acts or, more generally, to states of affairs. But one can also have attitudes to one’s acts or choices themselves, and similar questions arise about these attitudes’ rationality: if one has chosen the act producing the greater of two goods, should one regret not having chosen the other act? One kind of regret is not possible in this situation. One cannot regret not having chosen the act that would have been all things considered right, since one’s alternative act would have been all things considered

Whenever a state of affairs is intrinsically good, it is appropriate, rational, and perhaps good² to love that state for its own sake. By “loving” a state I mean being positively oriented toward it in desire, action, or feeling. This has three main forms: you can love a good state by desiring or wishing for it when it does not obtain, by actively pursuing it to make it obtain, or by taking pleasure in it when it obtains. Which specific form of love is appropriate to a particular good depends on facts about that good (does it obtain or not?) and about yourself (can you effectively pursue the good or not?). But for any intrinsic good, some positive orientation toward it, or some form of loving it, for its own sake is rational.

Unfortunately, we humans cannot love all good things with infinite intensity. We have finite capacities for desiring, actively pursuing, and taking pleasure in what is good, and the issue therefore arises of how it is best to divide our love between different good objects. Nor is the issue only one of our finite capacities. Even when we could love two goods more than we do, we can ask whether our existing love is divided between them in the best possible way.

Our initial examples raise this issue about division in a restricted context: where we could have produced either of two goods and chose one rather than the other. But we can also ask how much we should desire or wish for two goods neither of which now obtains, both when we can produce these goods through action and when we cannot. We can also ask how much pleasure we should take in two goods that do obtain. To all these questions the concentration and proportionality views give competing answers. If two goods obtain, for example, one greater than the other, the concentration view says it is best to take as much pleasure as possible in the greater good, even if this means taking no pleasure in the lesser good. The proportionality view says we should take more pleasure in the greater good but still some in the lesser good.

Although I find the concentration view coherent and intelligible, I do not believe that it is in the end credible. I find it more attractive to hold, with the proportionality view, that it is best to divide one’s love proportionally between different good objects. And several contemporary philosophers have endorsed this view’s implication for the special context following a choice between goods. Bernard Williams, Ronald de Sousa, Susan Hurley, Michael Stocker, John Kekes, Jonathan Dancy, and others have argued that when you have chosen a greater over a lesser good, it is rational to feel

wrong. But one can regret not having chosen an act with a certain right-making property, namely, that of producing what in these circumstances happens to be a lesser good. An analogue of the concentration view says this kind of regret is not rational; an analogue of the proportionality view says that it is. And of course there can be other right-making properties not connected with the production of good states of affairs. Although issues about attitudes to outcomes of acts and issues about attitudes to acts themselves are closely connected, I will confine my discussion in this article to the former issues. This will involve some narrowing of focus from the full range of cases where regret can be rational but will allow issues about monism and pluralism to arise in a clear and simple way.

² See Thomas Hurka, “Virtue as Loving the Good,” *Social Philosophy and Policy* 9 (1992): 149–68.

some regret for the forgone lesser good.³ They have found this claim intuitively plausible in itself, as I agree that it is. But the claim becomes even more persuasive when it is seen as but one application of the more general proportionality view. If two goods do not obtain, many will agree that it is rational for you to desire or wish for both, the greater more than the lesser but each to some degree. And if two goods do obtain, many will agree that it is rational for you to take pleasure in both. The claim about rational regret concerns a case that is in a way the overlap of these two. If you have chosen between two goods, one obtains while the other does not. Because of this, the forms of love appropriate to the goods are different: pleasure that it obtains for the one and regret or a wish that it obtained for the other. But, as in the earlier cases, it is intuitively plausible that you should feel some love for both goods and should divide this love in proportion to their degrees of goodness.

I have claimed that the more general proportionality view increases the plausibility of the claim about rational regret; at the same time, the view helps place limits on that regret. If you have just chosen one holiday from among ten possible holidays, you ought rationally to regret the holidays you will not take, but you should not do so to excess. You should not spend all your time away intensely regretting the nine trips you are not taking. The proportionality view can help explain why this is so.

Much recent writing about regret asks only whether regret after a choice is rational at all, without raising the further question of how much regret is rational. The proportionality view builds an answer to this second question into its answer to the first. If the reason regret is rational is that it involves a proportional division of love, then it is rational only when its intensity is proportioned to the value of its object. In particular, it is not rational to regret a forgone lesser good more intensely than one is pleased by a greater good one did produce.

This limit on rational regret, although helpful, is not sufficient. After all, each of the nine holidays you did not choose may be only slightly less good than the one you will take; if the regrets appropriate to these holidays are added together, the result is still an excess of negative feeling. But there are further reasons for limiting the intensity of your regret. One such reason is instrumental. If the good to be had from your holiday is pleasure, and intense regret would interfere with this pleasure, then the regret is instrumentally bad and to be restrained because of its effects. And there is a second, noninstrumental reason for limiting the kind of regret that involves pain at the absence of a good. Although this pain is a rational response to its object, it is also an instance of something evil—namely, pain—and should for that reason be minimized. The strength of these two reasons depends on how much

³ Bernard Williams, "Ethical Consistency," in his *Problems of the Self* (Cambridge: Cambridge University Press, 1973), pp. 172–75; Ronald de Sousa, "The Good and the True," *Mind* 83 (1974): 534–51; Susan Hurley, *Natural Reasons* (New York: Oxford University Press, 1989), pp. 171–74; Michael Stocker, *Plural and Conflicting Values* (Oxford: Clarendon Press, 1990), pp. 241–77; John Kekes, *The Morality of Pluralism* (Princeton, NJ: Princeton University Press, 1993), pp. 57–58; and Jonathan Dancy, *Moral Reasons* (Oxford: Blackwell, 1993), pp. 120–23.

weight we give the having of appropriate attitudes as against the positive and negative values of pleasure and pain, but we often seem to treat the reasons as substantial. Thus, we often do not tell friends bad news about ourselves to spare them the pain of sympathizing with us. Although sympathy would be a rational and even virtuous response on their part, we seem more concerned to spare them upset.

Finally, there is a limit on rational regret that follows from a plausible addition to the proportionality view. A simple view would say that the degree of love appropriate to a good depends only on that good's magnitude. If one good is greater than another, then, whatever else is true of them, one should care more about the first than about the second. But this view implies, implausibly, that we should have intense attitudes to what are only very remote possibilities. Imagine that you have just returned from a holiday marred by unseasonably bad weather. It is appropriate for you to regret missing the extra pleasure you would have had, given normal weather. But the simple view says you should regret more intensely the even greater pleasure you would have enjoyed had a stranger given you a million dollars on the beach or had aliens abducted you and taken you to an intergalactic pleasure palace. Or consider a similar example involving evils. If your child has just missed being struck by a car, you should feel relief that she did not suffer the pain of an accident. The simple view implies that you should feel more intense relief that she was not abducted by aliens and taken to an intergalactic torture chamber.

Because these implications are implausible, the proportionality view should be supplemented by a modal condition: when a good does not obtain, the degree of love appropriate to it depends on the degree to which its obtaining is a close possibility. If a good could exist now given only small changes in the world, then as much or almost as much love is appropriate to it as to an existing good of the same magnitude. As the possibility of its obtaining becomes more remote, however, the intensity of love appropriate to it diminishes. Beyond a threshold of remoteness perhaps no love is rational, even for a very great good.

This modal condition further limits the intensity of rational regret. Immediately after a choice a forgone good is a very close possibility, since it would obtain now had you only chosen differently a moment ago. As time passes, however, and the effects of your choice multiply, the changes in the world required to make that good actual become greater and, in consequence, the degree of love appropriate to it diminishes. If you have just chosen a holiday, it may be rational for you to feel as much regret for the trips you will not take as anticipation for the one that lies ahead. But as you reach your destination and accumulate experiences there, the possibility of being somewhere else becomes progressively more remote and progressively less an object of rational concern. It is not that you should regret something in the past less because it is in the past. On the contrary, what you should regret less is in part that you are not on a different holiday now, and you should regret this fact less because it now has less of a modal property that is necessary for objects at any time to merit serious love. Regret about a holiday, although rational, should be limited not only for reasons connected with pleasure

and pain but also because, over time, its intrinsic appropriateness diminishes. The regret is rational as an instance of proportional love, but like all such love it becomes less rational for more remote possibilities.

II. Monism and Rational Regret

I have argued that when you have chosen a greater over a lesser good, it is rational to feel some regret for the forgone lesser good. But many philosophers who accept this claim tie it to a further one that I believe is mistaken. De Sousa, Stocker, Kekes, and Dancy, and also David Wiggins and Martha Nussbaum, hold that regret for a lesser good is rational only when that good is an instance of a different generic good than is the greater, and is therefore rational only given a pluralistic rather than a monistic theory of what is good. A pluralistic theory contains several generic goods rather than only a single one—for example, pleasure, knowledge, and achievement rather than only pleasure. A monistic theory contains just one generic good. According to both monism and pluralism there are different individual good states of affairs. But whereas monism says there is a single explanation of these good states' goodness, or a single good-making property, pluralism says there are different explanations and different good-making properties. And many philosophers argue that there must be different good-making properties for regret for a lesser good to be rational.

I will argue, against this widespread view, that monism, too, allows rational regret. One can combine the proportionality view with a pluralistic theory of what is good, but equally well with a monistic theory. Issues about the division of appropriate attitudes and about the number of good-making properties are logically independent. To see this, let us examine the arguments commonly given for the claim that rational regret requires plural values.

Wiggins and Nussbaum advance this claim in the course of what they think is the closely related argument that only pluralism can account for the occurrence of weakness of will. If there are different good-making properties, they argue, we can explain why people sometimes fail to perform the act they judge to be best; given just a single good-making property, we could not.⁴ But, in fact, arguments about

⁴ David Wiggins, "Weakness of Will, Commensurability, and the Objects of Deliberation and Desire," in A. O. Rorty, ed., *Essays on Aristotle's Ethics* (Berkeley: University of California Press, 1980), pp. 241–66; and Martha C. Nussbaum, *The Fragility of Goodness* (Cambridge: Cambridge University Press, 1986), pp. 106–17. Like several other writers on this topic, Wiggins and Nussbaum speak not just of plural but of "incommensurable" goods. But the word "incommensurable" only creates confusion here. In some philosophers' usage, goods are incommensurable whenever they cannot be reduced to a single underlying value; in this sense of "incommensurable," which does not go beyond "plural," incommensurable goods can be fully and precisely comparable. In a stronger sense, goods are incommensurable when they cannot be compared. This stronger sense is irrelevant to the subject of this article, which concerns the rationality of regret for what is recognized to be a lesser good.

weakness of will and about rational regret are not closely related; what is more, we do not need plural values to account for weakness of will.

Weakness of will is a more extreme phenomenon than proportional regret. When we feel proportional regret, we have some love for the lesser of two goods but less than we do for the greater good. When we act weakly, by contrast, we love the lesser of two goods more, since we pursue it in preference to the greater good. Because of this, weakness of will is not on any view a rational response to two goods. On the contrary, it is clearly irrational to love a lesser good more than a greater. How then can an argument about weakness of will support a conclusion about rational regret? How can the conditions for one irrational phenomenon bear on those for the rationality of a different phenomenon?

Nor is it true that only pluralism can account for weakness of will. Imagine that you have a choice between instances of the same good—for example, between a greater and a lesser pleasure. You can weakly pursue the lesser pleasure if you are attracted to some property it has that does not bear on its degree of goodness. Not all human desires are for objects thought of as good; some are directed at objects independently of their degree of goodness. And weak-willed action can result when one of these desires is more powerful than any derived from a judgment of goodness. In a choice between pleasures, for example, the lesser pleasure may be closer in time, and although you do not believe its temporal location affects the value of a pleasure, you may nonetheless have within you a strong preference for closer pleasures, one that makes you weakly pursue what you acknowledge is a lesser good.⁵

Weakness of will, then, can be explained in terms of desires that are independent of judgments of goodness. But desires of this kind cannot figure in proportional regret, because the regret appropriate to a forgone lesser good is regret for it as a lesser good or as having some good-making property to a lesser degree. The regret is not independent of but reflects or parallels a judgment of lesser goodness. This fact underlies what is probably the most common argument for the claim that rational regret requires plural values, one given by, among others, de Sousa, Stocker, Kekes, and Dancy.

If you have chosen a more valuable instance of one generic good over a less valuable instance of another, this argument runs—for example, some more valuable pleasure over some less valuable knowledge—we can see the appropriateness of regret for the lesser good. The outcome of your act, although the best possible in your situation, lacks something present in the forgone outcome—namely, the distinctive value of knowledge—and you can rationally regret the absence of this value. But there is no similar lack if your choice was between instances of the same good, for example, between a greater and a lesser pleasure. Then everything present in the lesser good is present in the greater good, and there is no rational

⁵ There are similar points about weakness of will in Stocker, *Plural and Conflicting Values*, pp. 211–40.

ground for negative feeling. Stocker states this argument as follows. In a choice between instances of the same good,

there is no ground of rational conflict because the better option lacks nothing that would be made good by the lesser. Correlatively, the lesser good is not good in any way that the better is not also at least as good. There is no way, then, that the lesser option is better than the better one. And thus, there is no rational reason to regret doing the better—i.e. to regret doing it rather than the lesser.

Monism thus cannot allow for the lack and loss involved in rational conflict over practicable options.⁶

This argument clearly applies to some cases of choice between instances of the same generic good—namely, ones where the better instance includes the less good as a proper part. Imagine that you can give some person either five units of pleasure by playing her a certain song or ten units of pleasure by playing her first that song and then another. And imagine that you give her the ten units of pleasure. You cannot rationally regret not having produced the five units of pleasure, because you did produce them. Since the person got those very five units of pleasure, from the same song at the same time, the better outcome lacks nothing that is present in the less good.

Cases of this kind—let us call them “inclusion cases”—do not arise only for choices between instances of a single good. You can equally well have a choice between giving a person five units of pleasure and giving her the same five units of pleasure plus five units of knowledge. Here again, if you choose rightly and produce both the pleasure and the knowledge, you cannot rationally regret not having produced the pleasure, because you did produce it.

More important, not all cases of choice between instances of the same good are inclusion cases. Imagine, to take the simplest example, that you have a choice between giving five units of pleasure to one person, *A*, and giving ten units of pleasure to a different person, *B*. Here the lesser good is not included in the greater good as a proper part; it is not the case that if *B* enjoys the ten units of pleasure, *A* will enjoy the five units. Given this, it can surely be rational for you, if you produce the ten units for *B*, to feel some regret at not having produced the five units for *A*. There is, on the face of it, only one generic good at issue in your choice, namely, pleasure. But if the choice is between the pleasures of different people, it is not true, in Stocker’s words, that “the better option lacks nothing that would be made good by the lesser.” On the contrary, it lacks any pleasure for *A*. Nor is it true that “there is no way . . . that the lesser option is better than the better one.” If you have chosen a greater instance of one good for one person over a lesser instance of the same good

⁶ Stocker, *Plural and Conflicting Values*, p. 272. Stocker’s restriction to “practicable options” is intended to exclude regret about one’s choice situation, i.e., regret that one had to choose between two goods rather than being able to produce both simultaneously.

for another person, you can rationally regret not having produced what would have been better for the second person.

The distinction between inclusion and noninclusion cases shows that regret for a forgone lesser good requires this good to be distinct from the greater good. But there are two senses in which we can speak of one good's being "distinct" from another. In a weaker sense, *X* is a distinct good from *Y* if *X* is both an instance of some generic good and in some way distinct from *Y*. In a stronger sense, *X* is a distinct good from *Y* if it is an instance of a distinct generic good from *Y*. If individual good states of affairs are distinct in the stronger sense, they have different good-making properties; if they are distinct in the weaker sense, they need not have different good-making properties. In inclusion cases the forgone lesser good is not distinct from the greater good in even the weaker sense, and proportional regret is therefore not possible. But rational regret does not require the lesser good to be distinct in the stronger sense. It is enough if it is a distinguishable instance of the same generic good, that is, if it is distinct in the weaker sense. And it is distinct in the weaker sense if it is an instance of the same generic good in a different person.

Stocker acknowledges that we can rationally feel regret after choosing more happiness for one person over less happiness for another person. But he argues that the happinesses of the two people are distinct generic goods—that in the choice just described “this-person’s-happiness and that-person’s-happiness involve plural values: the differently-owned happinesses.”⁷ Now, one can perhaps hold that different people’s happinesses are distinct generic goods, although I do not find this very plausible (see sec. III below). But it is surely not necessary to do so. A monist who values only happiness can say that although the happinesses of different people are different individual goods, what makes them good is in each case the same—namely, just their involving happiness. Since the different happinesses all share the same good-making property, he can say, they are instances of the same generic good. And the monist can underscore his monism by applying his view to other cases. Thus, he can insist that a situation where *A* has zero units of happiness and *B* has fifteen is no less good than—it is just as good as—a situation where *A* has five units and *B* has ten, because it contains the same quantity of happiness. Since the two situations involve just the one generic good of happiness, which is present to the same degree in both, they are exactly equal in value. Despite this, the monist can say, if the zero/fifteen situation obtains, one can rationally regret that it contains no happiness for *A*.

It may be objected that what I am calling a monistic theory is so only in name. Precisely because it allows regret for the happinesses of different people, it is really a form of pluralism.⁸ But what is the basis for this objection? If monism is defined in the standard way, as a view about the number of good-making properties, then the theory I have described is clearly monistic. This is shown by its evaluations of

⁷ Stocker, *Plural and Conflicting Values*, p. 248.

⁸ This objection was raised against earlier versions of the article by Michael Stocker, Ruth Chang, and Tracy Isaacs.

the zero/fifteen and five/ten situations, which consider only the quantities of happiness the two situations contain and no other facts about them. Perhaps the objection assumes that a theory's allowing regret is itself sufficient to make the theory pluralistic. If so, however, the objection simply begs the question against my argument and turns its own conclusion into an empty tautology. The claim that only pluralism allows rational regret is surely meant to be a substantive one rather than one that follows trivially from the meanings of "monism" and "pluralism." But on this last understanding of the objection, the claim is just trivial. If allowing regret is definitionally sufficient for pluralism, the claim that only pluralism allows regret says no more than that a theory that allows regret must allow regret. And surely no one has meant to defend only this empty claim.

If monism is defined standardly, then, it can allow regret for a forgone lesser good when it is a good of a different person. It can in a similar way allow regret for a forgone lesser good of the same person at a different time. If you have chosen a greater pleasure today over a lesser pleasure tomorrow, you can rationally regret the forgone lesser pleasure, because its different temporal location makes it distinct in the weaker sense. But cases involving goods of the same person at the same time raise more complex issues.

Some philosophers argue that monism precludes rational regret by using examples of one person's choice between quantities of the same good at the same time. Thus, de Sousa asks us to imagine a choice between different quantities of money. If you have chosen, say, \$1,000 over \$500, what ground do you have to regret that you did not choose the \$500? Nussbaum has a similar example involving bagels. On one plate are two buttered bagels, and on another plate just one bagel. If the bagels are indistinguishable in their taste, texture, and so on, and you choose the two bagels, how can you rationally regret not having chosen the one?⁹

These examples are meant to show that there is no ground for regret after a choice involving just one intrinsic good. But they cannot show this, because the objects they involve are only instrumentally good. This is clearest of the money in de Sousa's example. Money is good not in itself but only as a means to the pleasure, knowledge, and other intrinsic goods it can help its owners pursue. The same is true of Nussbaum's bagels. The existence of a bagel has no value in itself, apart from the use people will make of it. Bagels too, therefore, are good only instrumentally, or as means to pleasures and other benefits for people. And on no credible view is it rational to feel regret for a forgone lesser instrumental good. If you choose to produce an intrinsic good by a more effective rather than a less effective means, you have no reason to bemoan your omission of the less effective means. Although regret is appropriate for forgone intrinsic goods, it is not so for forgone instrumental goods such as money and bagels. Because of this, examples involving instrumental goods show nothing about the rationality of regret for intrinsic goods.

⁹ De Sousa, "The Good and the True," pp. 536, 548; Nussbaum, *The Fragility of Goodness*, pp. 115–16. Stocker gives a similar example involving paint (*Plural and Conflicting Values*, p. 245).

Nonetheless, these examples may be used to raise an objection to my account of monistic regret. Imagine that in Nussbaum's example you choose the two bagels and enjoy the greater pleasure they offer. The lesser pleasure you would have derived from the other bagel, it may be argued, is distinct from any pleasure you did get, because it would have had a different cause. Since the forgone lesser pleasure would have had a distinct causal origin, it is not included in the greater pleasure. And this means that if distinctness in the weaker sense is sufficient for rational regret, you should regret not having had that lesser pleasure. But this is absurd: in your situation you have no ground for any regret, either for the forgone bagel or for the lesser pleasure it would have produced.

I agree that a plausible monism must deny that in Nussbaum's example you should regret the pleasure of the forgone bagel. It can do this, however, if it revises slightly its account of the conditions for rational regret and, in particular, narrows its definition of distinctness in the weaker sense.

According to our original definition, X is in the weaker sense a distinct good from Y if X is an instance of some generic good and in some way distinct from Y . We could handle Nussbaum's example by narrowing this definition so that a distinct good must be distinct in some way other than its causal origin, but I prefer a revision that is more broadly based and less ad hoc. On this revised definition, to count as distinct in the weaker sense, X must be distinct from Y in its intrinsic properties—that is, distinct apart from its relations, including its causal relations, to other states. To be weakly distinct from Y , X must differ from Y in its internal and not just in its external features. By this criterion, the pleasure you forgo in Nussbaum's example is not distinct from the pleasure you enjoy, since it does not differ from it internally. It involves all the same tastes, textures, and smells, only with a different causal origin. And this means that, on the revised view, you should not regret the forgone pleasure. Since it is not intrinsically distinct from the one you enjoy, it is not distinct in the way that matters for rational regret. A choice between pleasures from indistinguishable bagels is not literally an inclusion case, but for purposes of regret it is effectively equivalent to one.

This narrower definition of "distinct good" is appealing, since it not only yields the right result about Nussbaum's example but does so in a principled way. The distinction between intrinsic and nonintrinsic properties is a metaphysically important distinction and one that is already important for theories of the good. On a common view a state's good-making properties, the ones that make it intrinsically good, must themselves be intrinsic properties;¹⁰ if we accept this view, any regret appropriate to the state must be regret for it as having those intrinsic

¹⁰ See G. E. Moore, "The Conception of Intrinsic Value," in his *Philosophical Studies* (London: Routledge & Kegan Paul, 1922), pp. 253–76; Christine M. Korsgaard, "Two Distinctions in Goodness," *Philosophical Review* 92 (1983): 169–95; R. M. Chisholm, *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986); and Noah M. Lemos, *Intrinsic Value: Concept and Warrant* (Cambridge: Cambridge University Press, 1994).

good-making properties. But the regret must also be for it as a distinct instance of those good-making properties, and although there is no logical compulsion here, it is attractive to insist that the distinctness must rest on others among the state's intrinsic properties. On this view the intrinsic/nonintrinsic distinction plays two roles in value theory: it limits both the kinds of properties that can be good-making properties and those among a state's non-good-making properties that can distinguish it for purposes of rational regret. Distinctness in intrinsic goods, it is appealing to say, must be intrinsic distinctness.

It may be objected that this narrower definition of weak distinctness is inconsistent with my earlier treatment of regret for forgone goods of different people and at different times. Is its being a good of this person or at this time not a relational property of a good and therefore excluded by the current proposal from grounding rational regret?¹¹ I do not believe that it is. If there is a relation in different-person and different-time cases, it is not a relation of the good state to other states, and it is also not an external relation. States such as pleasures do not exist in free-floating independence, with only accidental attachments to persons and times. Each is essentially and internally a state of this person at this time and distinguished from other pleasures in part by these facts. What is more, the time and ownership of a pleasure can, like its other intrinsic properties, be detected by introspection. Each person can tell introspectively whether a certain pleasure is one she is feeling now. The best description of the generic good of pleasure is not "there being pleasure" but "some person's feeling pleasure at some time." When pleasure is described in this way, its instances have particular persons and times not as external relata but as intrinsic aspects or features.

Let us assume, then, a monistic theory that accepts this narrower definition of distinctness in the weaker sense. This theory can hold that it is often rational to feel regret after choosing between instances of the same good for the same person at the same time. Exactly when regret is rational, however, can vary between generic goods and even given different understandings of the same good.

Consider, first, knowledge as a generic good. If you have chosen some more valuable scientific knowledge over some less valuable historical knowledge, you can rationally regret the forgone knowledge of history because its different subject matter makes it distinct in the weaker sense. Since its propositional content is an intrinsic feature of every item of knowledge, regret is appropriate after any choice concerning knowledge except a choice to acquire knowledge of the same proposition by a different means, say, by reading a book rather than hearing a lecture.

Now consider pleasure as a good. Given the revised definition of distinctness, a monistic theory that values only pleasure will agree that it is not rational to regret forgone pleasure from a qualitatively identical bagel. But what the theory says about other cases involving the same person's pleasures at the same time depends on its choice between two different views of what pleasure is.

¹¹ I owe this objection to Ruth Chang.

Let us assume, as a starting point, that pleasures are sensations or feelings distinguished by an introspectible property of pleasantness. One view, which I will call the Benthamite view, says that pleasure sensations have the introspectible property of pleasantness and no other introspectible properties. These sensations are, as it were, atoms of pleasantness, with no further internal features to distinguish them from each other. Pleasures from different sources, such as those of eating bagels and discussing philosophy, do not on the Benthamite view differ from each other intrinsically. If we experienced both pleasures without knowing their causal origins, we could not tell them apart, since they both have just the one intrinsic property of pleasantness.

A different and, in my opinion, more plausible view says that sensations with the property of pleasantness always have other introspectible properties that can differ from one such sensation to another. On this non-Benthamite view, we never experience pleasantness on its own but always in conjunction with other introspectible properties. Shelly Kagan has expounded this view using a helpful analogy with the loudness of sounds.¹² Loudness is a property of sounds, and one by which they can be ranked. But it is impossible to experience loudness on its own, apart from a certain pitch, timbre, and so on. Because of this, loudness in a tuba is qualitatively different from loudness in a flute. The situation is similar, the non-Benthamite view claims, with the pleasantness of sensations. The pleasures of eating bagels and of discussing philosophy share the common property of pleasantness, but each has many other introspectible properties that make them intrinsically very unlike.

Imagine a monistic hedonism that accepts the first, Benthamite, view of pleasure. If you have chosen a greater pleasure of eating bagels over a lesser pleasure of discussing philosophy, this theory will say you have nothing to regret in the forgone lesser pleasure, since it is not intrinsically distinct from the pleasure you did enjoy. You may rationally regret that you had to choose between the two pleasures rather than being able to enjoy both simultaneously, but this kind of regret, regret about your situation, can be appropriate even given the concentration view. And if we take your situation as given, the Benthamite view does not allow you rationally to regret that you did not enjoy the lesser pleasure rather than the greater. By contrast, the more plausible, non-Benthamite view does allow you to feel regret. On the non-Benthamite view, the two pleasures available to you are distinct in the weaker sense, and while enjoying the greater pleasure of eating bagels, you can rationally regret missing out on the qualitatively distinguishable lesser pleasure of discussing philosophy. You can appropriately regret a forgone pleasure at the same time, because its different origin gives it different internal properties.

Stocker agrees that pleasures can differ introspectibly, contrasting eloquently the pleasure of a “hot and smooth” omelette with that of a “cold and crunchy”

¹² Shelly Kagan, “The Limits of Well-Being,” *Social Philosophy and Policy* 9 (1992): 172–73.

salad and “piquant” pleasures with “languorous” ones.¹³ But he seems to assume that pleasures distinct in this weaker sense must also be distinct in the stronger sense of being instances of different generic goods. According to his theory, just as the pleasures of different people are different generic goods, so are pleasures with different introspectible properties.¹⁴ As before, however, although one can perhaps hold this pluralistic view, one need not. A monistic hedonism that accepts the non-Benthamite view can agree that the pleasures of eating bagels and of discussing philosophy differ introspectibly, perhaps in many ways. It can also agree that it is rational to feel regret after choosing between these pleasures. But it can say that what makes the pleasures good is in each case the same—namely, just their being pleasant. And it can underscore its monism by insisting that, because it involves the same quantity of pleasure, a situation where you get zero units of pleasure from discussing philosophy and fifteen units from eating bagels is just as good as one where you get five units from discussing philosophy and ten units from eating bagels. Although the two pleasures in these situations differ intrinsically, which allows rational regret, they share the same one good-making property.¹⁵

The distinction between the Benthamite and non-Benthamite views of pleasure explains when monistic hedonism will and will not allow rational regret, and it may also have a broader significance. Many philosophers who argue that monism precludes rational regret use hedonism as their model of a monistic theory, but they apply hedonism in a very restricted way. They consider mostly or only choices between one person’s pleasures at one time, and often tacitly associate hedonism with the Benthamite view. This combination of assumptions may explain why these philosophers tie proportional regret to pluralism, since in the situation that the assumptions define, regret is indeed not rational. As I have argued, however, the assumptions do not generalize to all choices between instances of the same good. Even a Benthamite hedonist can rationally feel regret after choosing a greater pleasure for one person over a lesser pleasure for another, or a greater pleasure at one time over a lesser pleasure at another. What is more, many versions of hedonism, including the most plausible, are not Benthamite and therefore also allow regret after a choice between pleasures from different sources. Finally, no analogue of the Benthamite view is plausible for generic goods other than pleasure. Instances of knowledge and of other candidate goods such as achievement differ intrinsically

¹³ Stocker, *Plural and Conflicting Values*, pp. 190, 267–68.

¹⁴ See esp. Stocker, *Plural and Conflicting Values*, pp. 267–68.

¹⁵ As before, it may be objected that to accept the non-Benthamite view and allow regret for introspectibly different pleasures is to abandon monism in favor of something equivalent to pluralism. But, again, what is the basis for this objection? The theory I have described holds that there is only one good-making property—i.e., pleasantness—and underscores its monism in the way it evaluates the zero/fifteen and five/ten situations. And if the objection treats allowing regret as itself a sufficient sign of pluralism, it, again, simply begs the question against my argument and reduces its own conclusion to an empty tautology.

in ways that in most real situations allow for rational regret. In a small range of choices defined partly by Benthamite assumptions, monism does indeed not allow rational regret; outside that range, however, it is as compatible with proportional attitudes as is pluralism.

III. Moderate Versus Extreme Pluralism

A reader may grant the argument of the previous section but question its significance. The claim I have challenged—that monism precludes rational regret—is often used by philosophers who want to defend pluralism. They take regret for a forgone lesser good to be clearly rational and argue that if only pluralism allows this regret, we should prefer pluralism to any version of monism. That monism, too, allows rational regret shows that this particular argument is unsuccessful. But this conclusion, a reader may say, is of limited interest, since there are many other reasons why a plausible theory of the good must be pluralistic. Even if sound, my argument provides at best a minor defense of what is on other grounds a thoroughly implausible view of the good.

This article is not the place to discuss the overall merits of monism and pluralism, so let me grant for argument's sake that the most plausible theories of the good are pluralistic. Even so, my argument is important, because it affects in several ways our understanding of pluralism.

Consider, first, a pluralistic theory with three generic goods: pleasure, knowledge, and achievement. This theory allows choices not only between instances of different goods (for example, between a pleasure and an item of knowledge), but also between instances of the same good (for example, between a greater and a lesser pleasure). If rational regret required plural values, this pluralistic theory could call for regret only after the first kind of choice but not after the second; if there is no such requirement, it can call for regret after both kinds of choices. Even assuming pluralism, then, my argument extends the number of cases in which regret for a lesser good is rational.

Second, the argument favors what I will call moderate over extreme pluralism. A moderate pluralism contains just a moderate number of generic goods—say, three, ten, or twenty—whereas an extreme pluralism contains an immense number of generic goods. Stocker defends an extreme pluralism. As we have seen, he holds that the happiness or pleasure of every different person is a distinct generic good, as is every introspectibly distinguishable pleasure. The previous section argued only that this extreme pluralism, although perhaps possible, is not forced upon us. Now I will argue that, given this conclusion, we should reject extreme pluralism. Even if the most plausible theory of the good is pluralistic, it contains just a moderate and not an immense number of generic goods.

My argument will assume a minimal version of Ockham's razor: if two theories are equivalent in all respects except that one contains more generic goods, we

should prefer the simpler theory. Generic goods should not be posited unless doing so makes some further evaluative difference.

Given this assumption, let us ask how an extreme pluralism such as Stocker's evaluates the two pairs of zero/fifteen and five/ten situations discussed above. In the first pair, either *A* has zero units of pleasure and *B* has fifteen or *A* has five units and *B* has ten; in the second pair, one person, *A*, gets either zero units of pleasure from discussing philosophy and fifteen from eating bagels or five from discussing philosophy and ten from eating bagels. Monistic hedonism says the situations involve just the one generic good of pleasure, and since this good is present to the same degree in both, they are exactly equal in value. Stocker's extreme pluralism says they involve more than one generic good. Let us assume that it says they involve three: pleasure, pleasure for *A*, and pleasure for *B* in the first pair and pleasure, pleasure of discussing philosophy, and pleasure of eating bagels in the second.¹⁶ What does a theory with these generic goods say about the relative values of the two situations?

One possibility is that the theory agrees with monistic hedonism that the two situations are exactly equal in value. Although it must consider three generic goods rather than only one to reach this conclusion, it agrees that the zero/fifteen and five/ten situations are equally good and in fact agrees with monistic hedonism in all its evaluations of situations involving pleasure. In this case we should reject extreme pluralism, since its extra goods do not affect its judgments about states of affairs and are therefore idle. This conclusion would not follow if rational regret required plural values, for then the extra goods would make a difference to which attitudes are rational. But if one can rationally regret a forgone instance of the same generic good, as the previous section argued, an extreme pluralism that merely mimics the judgments of a simpler theory is pointless.

To be defensible, then, an extreme pluralism must make different judgments about states of affairs—for example, preferring the five/ten situations to the zero/fifteen. So, let us imagine making these judgments. This still does not justify extreme pluralism, for there may be other, simpler explanations for why the five/ten situations are better than the zero/fifteen.

Let us start with the second pair of situations, involving pleasures from different sources. We might think the situation with five units of the pleasure of discussing philosophy and ten of the pleasure of eating bagels is better because, pleasure aside, discussing philosophy is a more valuable activity than eating. This reason does not require treating every distinguishable pleasure as a distinct generic good; it merely requires adding intellectual activity or something like it to pleasure on a short list of goods. Or perhaps the social aspect of discussing philosophy makes it

¹⁶ A slightly different theory says the situations involve just two generic goods: pleasure for *A* and pleasure for *B* in the first pair and the pleasure of discussing philosophy and the pleasure of eating bagels in the second. (This theory denies that pleasure as such is a generic good.) My arguments against the three-goods view tell just as much against this slightly less extreme alternative.

better; this again requires just a moderate pluralism with social interaction as one extra good. To eliminate explanations of this kind, we must consider pleasures that do not differ in the nonhedonic goods that accompany them: say, the pleasure of eating poppy seed bagels and the pleasure of eating sesame seed bagels. Even here there may be a reason for preferring the five units of one pleasure and ten of the other. We may think that someone who has experienced both pleasures has more knowledge, both of bagels and of pleasure sensations, than if they had experienced only one. To eliminate this possibility, imagine that *A* has eaten many of both types of bagel and is fully knowledgeable about both. Given this condition, and considering the bagel pleasures only as pleasures, is the five/ten situation better than the zero/fifteen? I do not think many will now find this judgment plausible. And I especially do not think many will find plausible a further judgment that follows from it: that where *A* would get slightly more pleasure from eating only sesame seed bagels, *A* should nonetheless, and considering only the pleasures involved, prefer the mixture with some poppy seed bagels. The judgment needed to make pluralism about different pleasures defensible is not, when properly isolated, credible.

Now consider the first pair of situations, involving pleasures for different people. Here there is less temptation to be distracted by extraneous goods such as intellectual activity, and we can eliminate these goods by stipulating that the pleasures available to *A* and *B* come from the same source: they are all pleasures of discussing philosophy or pleasures of eating poppy seed bagels. Even given this stipulation, many will here make the judgment whose parallel did not seem plausible in the other pair of situations. Even considering *A*'s and *B*'s pleasures only as pleasures, many will say the five/ten situation is better than the zero/fifteen, because it distributes pleasure more equally. And they will also affirm the further judgment that a more equal distribution of pleasure can be better than a less equal one that contains more pleasure. Because these egalitarian judgments are attractive, I suspect many will find extreme pluralism about different people's pleasures more plausible than a similar pluralism about pleasures from different sources. But the acceptability of these judgments still does not justify this extreme pluralism, for, again, there may be simpler explanations for why more equal distributions of pleasure are better.

One such explanation appeals to two generic goods: pleasure and equality in the distribution of pleasure. According to this view, a full evaluation of the zero/fifteen and five/ten situations has two parts. First, it considers the total pleasure the situations involve; here the two come out equal. Then it considers the degree of equality they embody; here the five/ten situation is better, and this makes it all things considered better. The egalitarian judgment requires not a separate generic good for every person but just the one additional good of equality.

It may be objected that this allegedly moderate pluralism presupposes the extreme one, since equality among *X*'s cannot be a generic good unless the *X*'s themselves are distinct generic goods. But why should we accept this condition?

Even if the intrinsic goodness of equality among X 's requires the X 's to be intrinsic goods—which many would dispute—it does not follow that they must be distinct generically. Pleasures can be divided into classes according to the person whose pleasures they are. Why can we not value equality among these classes while holding that the pleasures themselves are good only as instances of pleasure?

A different explanation of the egalitarian judgment does not treat equality itself as good. Instead, it gives priority to increases in the pleasure of those who have less pleasure.¹⁷ According to one version of this view, each person's first unit of pleasure, either at a time or in the person's life as a whole, has a certain value, while every subsequent unit has progressively less value. On this view the five/ten distribution is better than the zero/fifteen, because A 's first five units of pleasure count for more than B 's eleventh to fifteenth units. Again, this explanation is simpler than one that treats the pleasures of different people as distinct generic goods. And if it is suggested that the explanation presupposes this more extreme pluralism, we can again ask why. Why can we not make a pleasure's degree of value depend on facts about a class to which it belongs while holding that what makes it good is just its being a pleasure?

Even if we prefer more equal distributions of pleasure, then, there are simpler explanations of this judgment than the one given by extreme pluralism. And this means that, given a minimal version of Ockham's razor, we should reject extreme pluralism as needlessly complex. As before, this argument would not be persuasive if rational regret required plural values, for then the extra goods in extreme pluralism would allow us rationally to feel regret after a choice between different people's pleasures. But if there is no such requirement, extreme pluralism complicates the theory of the good to no substantive effect. Even if the most plausible theory of the good is pluralistic, its form is that of moderate, and not extreme, pluralism.

IV. Conclusion

The claim that rational regret requires plural values has radical implications. Since regret can be appropriate after many choices—between the pleasures of different people or from differently flavored bagels—the claim requires us to posit an immense number of generic goods. Stocker, to his credit, sees this, but his extreme pluralism puts pressure on the concept of a generic good. What is the point of distinguishing between generic and individual goods if we almost never face choices between instances of the same generic good? And there is a simpler view

¹⁷ For the distinction between these two egalitarian views, see Dennis McKerlie, "Egalitarianism," *Dialogue* 23 (1984): 223–37; Larry S. Temkin, *Inequality* (New York: Oxford University Press, 1993), pp. 245–48; and Derek Parfit, "On Giving Priority to the Worse-Off" (ms., All Souls College, Oxford, 1989).

available: to allow that regret can be rational for intrinsically distinct instances of the same generic good, and then affirm no more than a moderate number of good-making properties.

Given its radical implications, we may wonder why so many philosophers have embraced the claim about regret and plural values. Let me offer a semiserious explanation. Much recent moral philosophy reads as if it were inspired by the following argument: "Utilitarianism is a false moral theory; therefore, the more unlike utilitarianism a moral view is, the more likely it is to be true." (The specifically British version of the argument runs: "Hare's is a false moral theory; therefore, the more unlike Hare's a moral view is, the more likely it is to be true.") To one who accepted this argument, extreme pluralism would indeed seem attractive. But, obviously, often the best replacement for a false theory corrects it only at some points and leaves others of its claims standing. The classical pluralistic theories of the good—for example, those of Moore, Rashdall, and Ross—usually agreed with hedonistic utilitarianism that pleasure is one intrinsic good and supplemented pleasure with only a moderate number of other generic goods. These theories are a more fruitful subject for study today than any fashionable extremes of antiutilitarianism.

How Great a Good Is Virtue?

Claims about the good, and especially about the human good, are standardly divided into two classes: *subjective* or *welfarist claims*, and *objective* or *perfectionist claims*. Welfarist claims make each person's good depend on certain of her subjective mental states, such as her pleasures or desires. Hedonism, which holds that only pleasure is intrinsically good, is a version of welfarism, as are views that equate a person's good with the fulfilment of her desires. Perfectionist claims, by contrast, hold that certain states of humans are good objectively, or independently of their connection to pleasures or desires. Thus, perfectionists have held that knowledge, achievement, and deep personal relations are intrinsically good regardless of how much a person wants or enjoys them, and that their absence impoverishes a life even if it is not a source of regret.

A common perfectionist claim is that moral virtue is intrinsically good and moral vice intrinsically evil. Benevolence, courage, and similar traits in themselves make a life better, and vices, such as malice, make it worse. This claim need not hold that virtue is the only intrinsic good. It can and in my view should recognize other perfectionist goods, such as knowledge, and even welfarist goods, such as pleasure. But it holds that among the intrinsic goods is moral virtue and among the evils is vice.

Given this perfectionist view, a natural question is how the value of virtue compares with that of other goods, such as pleasure and knowledge. Is virtue, as some claim, the greatest intrinsic good, even infinitely more valuable than pleasure or

I presented an earlier version of this chapter at the University of Maryland and of its last part at the University of Calgary and the Canadian Philosophical Association Annual Congress. For helpful comments I thank the audiences on those occasions, as well as Samuel Kerstein, Dennis McKerlie, Jamie Mayerfeld, Derek Parfit, Donald Regan, Michael Slote, Christine Swanton, Terry Teskey, and the students in my seminars at the University of Calgary. I am also grateful for support from the Social Sciences and Humanities Research Council of Canada.

knowledge? Is vice in a similar way the greatest evil? Or are virtue and vice each just one intrinsic value among others, sometimes outweighing other values and sometimes outweighed by them? I shall examine this comparative question. Against what I think is the most widespread view, I shall argue that, far from being the greatest good, virtue is a lesser good in the following sense: the value of a virtuous response to a good or evil object is always less than the value, either positive or negative, of that object. In a similar way, vice is a lesser evil. It is not that every instance of virtue or vice has less value than any other good or evil; on the contrary, it may outweigh many goods and evils. But it always has less value than the specific good or evil that is its object. Although important additions to the list of goods and evils, virtue and vice have in this sense a subordinate comparative status.

The question of virtue's comparative value can arise given any account of what virtue is, but it is most usefully discussed with a specific account in view. Let me therefore sketch an account which I find attractive and which also raises the comparative question especially sharply. This account treats virtue and vice as higher-level values in a specific way. I call it the *recursive account*, because it combines two elements: a recursive characterization of intrinsic goods and evils and a definition of virtue and vice that fits this characterization.¹

I. The Recursive Account

The recursive characterization starts by identifying certain states other than virtue and vice as intrinsically good and evil. It does not matter exactly what these states are, but let us assume the characterization starts with the following mixed welfarist-perfectionist base clauses about intrinsic goods and evils:

(BG) Pleasure, knowledge, and achievement are intrinsically good.

(BE) Pain, false belief, and failure are intrinsically evil.

These clauses, and in fact all of the recursive characterization, are agent neutral in form. Thus, (BG) says that person *A*'s pleasure is good not just from *A*'s point of view but from that of all moral agents, so all moral agents have the same reason to pursue this good of *A*. (BE) makes a similar claim about the evil of *A*'s pain.

The characterization then supplements these base clauses with four recursion clauses assigning intrinsic values to certain attitudes to intrinsic goods and evils.

¹ For earlier presentations of the elements of this account, see Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, trans. Roderick M. Chisholm and Elizabeth Schneewind (New York: Routledge, 1969); G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903); Hastings Rashdall, *The Theory of Good and Evil*, 2 volumes (Oxford: Oxford University Press, 1907); W. D. Ross, *The Right and the Good* (New York: Oxford University Press, 1930); Robert Nozick, *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981); Roderick M. Chisholm, *Brentano and Intrinsic Value* (New York: Cambridge University Press, 1986); and my "Virtue as Loving the Good," *Social Philosophy and Policy*, 9 (1992): 149–68.

The first recursion clause affirms the intrinsic goodness of loving what is good for itself:

(LG) If x is intrinsically good, loving x (desiring, pursuing, or taking pleasure in x) for itself is also intrinsically good.

By “loving” x , (LG) means having a positive attitude or orientation toward x . This has three main forms. One can love x by desiring or wishing for x when it does not obtain, by actively pursuing x to make it obtain, or by taking pleasure in x when it does obtain. By loving x “for itself,” (LG) means loving x apart from its consequences or for its own sake, that is, loving instances of x because they are instances of x . Thus, one loves pleasure for itself if one loves instances of pleasure because they are pleasures, and one loves knowledge for its own sake if one loves instances of knowledge because they are knowledge.

To see more concretely what (LG) implies, let us combine it with the base clause (BG), which says that a pleasure felt by person A is intrinsically good. (LG) adds that, if another person B loves A 's pleasure for itself—for example, if B desires or pursues A 's pleasure for itself—his doing so is also intrinsically good. Since (LG)'s application iterates or recurs, (LG) also says that, if a third person C loves B 's love of A 's pleasure for itself—for example, if C is pleased by B 's desire for A 's pleasure—this, too, is intrinsically good. To take another example, (BG) says that knowledge is intrinsically good. (LG) adds that pursuing knowledge for its own sake is intrinsically good, as is being pleased to be pursuing knowledge.

Parallel to (LG) is a second recursion clause affirming the intrinsic evil of loving for itself what is evil:

(LE) If x is intrinsically evil, loving x for itself is also intrinsically evil.

Combined with (BE), (LE) says that person B 's desiring or pursuing A 's pain for itself is intrinsically evil, as is B 's taking pleasure in A 's pain. If a third person C loves B 's love of A 's pain for itself—for example, by taking pleasure in B 's malicious pleasure—that, too, is intrinsically evil.

Two final recursion clauses concern the contrary attitudes of hating goods and evils:

(HG) If x is intrinsically good, hating x (desiring or pursuing x 's not obtaining, or being pained by x 's obtaining) for itself is intrinsically evil.

(HE) If x is intrinsically evil, hating x for itself is intrinsically good.

Combined with (BG), (HG) makes it intrinsically evil for B to desire or seek for itself the destruction of A 's pleasure or to be pained by A 's pleasure. Combined with (BE), (HE) makes it intrinsically good for B to be sympathetically pained by A 's pain—to feel compassion for A 's pain—or to desire or try to relieve it.

Behind the recursion clauses (LG)-(HE) is a comparatively simple idea: that appropriate attitudes to goods and evils are intrinsically good and inappropriate attitudes intrinsically evil. The appropriate attitude to an intrinsic good is to love it

for its own sake, and to an intrinsic evil is to hate it. The recursion clauses therefore make these attitudes intrinsically good and their contraries intrinsically evil.

On its own, the recursive characterization says nothing about moral virtue and vice. But this omission is easily remedied by adding a definition of the virtues and vices that fits the characterization:

(DV) The moral virtues are those attitudes to goods and evils which are intrinsically good, and the moral vices are those attitudes to goods and evils which are intrinsically evil.

Combined with (LG)-(HE), (DV) equates the virtues with loving intrinsic goods and hating intrinsic evils, and the vices with loving evils and hating goods. More specifically, (DV) equates the virtues and vices with all the higher-level goods and evils in a multilevel theory of value. The recursive characterization starts by identifying certain base-level goods and evils, namely, those of (BG) and (BE), and then adds an infinite series of higher-level goods and evils, each consisting in an appropriate or inappropriate attitude to a lower-level good or evil. (DV) equates the moral virtues with all the higher-level goods in this multilevel theory and moral vice with all the higher-level evils. The virtues and vices are intrinsic goods and evils that are made such by their orientation to other goods and evils.²

I find this recursive account of virtue and vice attractive. As my brief presentation of it has shown, it counts a benevolent desire for another's pleasure and compassion for his pain as virtues and malice toward him as a vice. When the account is elaborated and refined, it illuminates a wide range of further virtues and vices. The account fits naturally in a consequentialist framework, since it analyzes virtue and vice in terms of the central consequentialist properties of intrinsic goodness and evil. For those who accept consequentialism, therefore, it can give a complete account of virtue and vice. Even those who do not accept consequentialism can see it as giving a partial account of virtue and vice, one appropriate to the part of morality concerned with intrinsic goods and evils. They may want to supplement

² Unlike some definitions, (DV) finds virtue and vice in occurrent attitudes rather than in longer-lasting dispositions or traits of character. There are three reasons for preferring this approach. First, any disposition relevant to virtue can only be identified as a disposition to have certain occurrent attitudes—for example, as a disposition to have the desires, perform the actions, and feel in the ways valued by some part of the recursive characterization. Second, when it is considered apart from the occurrent attitudes in which it can issue, any such disposition has minimal value. As Aristotle says, the mere possession of a virtuous disposition, as in someone asleep or otherwise inactive, counts for little. Finally, though one can hold that occurrent attitudes have more value, or even have value only, when they issue from a virtuous disposition, this is in my view a mistake. Imagine that one person feels compassion that issues from a long-lasting disposition while another feels an otherwise identical compassion but only once, without its reflecting anything permanent in his character. If we set aside any (minimal) value the first person's disposition has in itself, and also set aside the values in the other feelings and actions to which this disposition will lead at other times, is the first person's compassion better than the second's? I see no reason to believe this, and therefore no reason to find dispositions specially important for virtue and vice as intrinsic values.

the account with further claims about nonconsequentialist forms of virtue, but can accept it as analyzing an important range of virtues and vices.

The account also raises immediately the question about the comparative values of virtue and vice. The account implies that virtue and vice cannot be the only intrinsic values; their goodness requires, by definition, other values to which they can be intentionally related. So how do the values of virtue and vice compare with those of base-level goods and evils, such as pleasure, pain, and knowledge? How great a good is virtue, and how great an evil is vice?

II. Virtue as a Lesser Good

Some philosophers hold, high-mindedly, that virtue is the greatest intrinsic good. The strongest such view holds that virtue is infinitely greater than other goods, or has lexicographic priority over them. W. D. Ross affirms part of this view when he writes that “no amount of pleasure is equal to any amount of virtue...; in other words, that while pleasure is comparable in value with virtue (i.e., can be said to be less valuable than virtue) it is not commensurable with it, as a finite duration is not commensurable with infinite duration.”³ John Henry Cardinal Newman makes a similar claim about vice, saying it would be less evil for all humankind to die “in extremest agony” than that “one soul...should commit one venial sin.”⁴ According to the fully general version of this lexicographic view, even the most trivial instance of virtue or vice outweighs the greatest imaginable quantity of any base-level value, such as pleasure, pain, or knowledge.

But the lexicographic view has extremely implausible implications. Consider two possible worlds.⁵ In the first world, natural conditions are benign and people therefore lead very pleasant lives, but they are all entirely self-concerned, caring about no one’s enjoyment but their own. In the second world, conditions are less favorable, which gives people less pleasure, but they are benevolent and sympathetic toward each other. If we believe that virtue is intrinsically good, we shall say the second world can be better than the first, as seems right to me. But the lexicographic view implies that the second world is better even if it contains vastly less pleasure—its inhabitants suffer agony while those in the first world are blissful—and only slightly more virtue—one person feels a tiny bit more compassion. Ross and Newman are committed to accepting this implication, but I do not see how they can. Or consider a single life, that of a virtuous person who suffers

³ *The Right and the Good*, p. 150. See also Ross, *The Foundations of Ethics* (New York: Oxford University Press, 1939), p. 275. He makes a similar claim about the infinite superiority of virtue over knowledge, his other main good, in *The Right and the Good*, p. 152.

⁴ *Certain Difficulties Felt by Anglicans in Catholic Teaching*, vol. 1 (London: Longmans, 1901), p. 240; quoted in Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984), p. 49.

⁵ Compare *The Right and the Good*, pp. 134–35.

agonizing pain. The lexicographic view implies that this life is overwhelmingly good with only an insignificant element of evil. But surely most of us think this life is on balance undesirable, one it would be better not to live.

The lexicographic view has equally implausible implications for cases involving action. As Hastings Rashdall asks in a penetrating critique, must we always direct our charitable contributions to people's moral improvement rather than to their education or material comfort, or at least do so when the two conflict? If the only way to save twenty innocent people from intensely painful torture is by bribing a corrupt official and thereby further damaging his character, is bribing the official wrong? Or consider a case involving just the agent's virtue. A person contemplating a career such as nursing or surgery may know that in it she will do much to relieve others' pain. But she may also rightly fear that the constant exposure to pain will harden her character, changing her from someone who feels deeply for others' pain to one who handles it efficiently but without accompanying emotion. If virtue had infinite comparative value, it would be wrong for this person to take up nursing or surgery whatever the benefits to other people. But surely that is absurd.⁶ It may be objected that an unemotional nurse can still act virtuously in her career, since she can still choose her work for its own sake or for the good it does to others. But if compassionate feeling is an additional form of virtue, as it surely is, and its loss would make her life on balance less virtuous, the lexicographic view still says her choosing the career would be wrong. We can imagine a more radical example in which the exposure to suffering hardens her character further, so she no longer chooses her work for others' sake but does it only as a job or for the money it provides. (Those familiar with careers such as nursing and surgery will know this example is anything but unrealistic.) In this case, and in fact if there is even the slightest probability of this more radical outcome, the lexicographic view says this person should not take up nursing whatever the benefits to others.

Despite rejecting the lexicographic view, Rashdall still holds that virtue is the greatest good: "It seems to me perfectly clear that the moral consciousness does pronounce some goods to be higher, or intrinsically better than others; and that at the head of these goods comes virtue."⁷ But it is not clear that this nonlexicographic view is possible. Consider two nonmoral goods, such as pleasure and knowledge. If neither is lexicographically prior to the other, some very intense pleasures outweigh some trivial items of knowledge and some very important knowledge outweighs some minor pleasures. But if some instances of each good outweigh some of the other, how can we say that either good is *in general* greater? If we have measured both goods on cardinal scales, we may have a formula for converting between these scales, and this formula may make, say, two units of pleasure equal to one of knowledge. But this 2:1 ratio depends entirely on our choice of units on the

⁶ *The Theory of Good and Evil*, vol. 2, pp. 41–47.

⁷ *The Theory of Good and Evil*, vol. 2, p. 37.

two scales, which is arbitrary. Given different units there could be a 2:1 ratio in the other direction. So if neither of two goods is lexicographically prior to the other, how can we make any comparison in the abstract between them?

This difficulty does not arise for the specific good of virtue, at least as understood by the recursive account. This is because every instance of virtue is, according to that account, specially connected to another good or evil that is its object. Every lowest-level virtue, to start, has as its object an instance of a base-level value, such as pleasure, pain, or knowledge. It is a desire for a pleasure, a pain at a pain, or something similar. We can therefore ask how the value of the virtue compares to that of its specific base-level object. If the virtue's value is always greater than its object's, there is a clear sense in which this form of virtue is a greater good than pleasure or knowledge; if its value is always smaller, the virtue is a lesser good. The same approach can be extended to higher-level virtues. Every such virtue is likewise connected to a specific base-level value, through the lower-level attitude that is its object. We can therefore assess the value of virtue as a whole by asking how in general the value of a virtuous attitude compares to that of its object. If the attitude's value is always finitely greater, there is a restricted but clear sense in which virtue is a nonlexicographically greater good. It may be that for every instance of virtue there are some instances of pleasure and knowledge that are better than it and some that are less good. But if there is a constant relationship between the value of the virtue and that of the specific base-level state to which it is connected, this relationship can underwrite a general claim about the comparative value of virtue. A similar relationship can underwrite claims about the comparative evil of vice. I think a charitable reading of Rashdall's nonlexicographic view takes it to rest on this kind of restricted comparison. He thinks virtue is a greater good and vice a greater evil because he thinks any virtuous or vicious attitude, though outweighed by some base-level values, has more goodness or evil than the specific base-level state that is its direct or indirect object.

Some philosophers argue that virtue is a greater good on the ground that virtue alone is morally good, whereas pleasure, knowledge, and the like are nonmorally good.⁸ (This is also sometimes given as a reason for the lexicographic view.) But this argument is not persuasive. If it assumes that moral goodness is a distinct property from ordinary intrinsic goodness, it cannot hold that virtue has only that property and no other. If we can compare the values of the two worlds described above, the pleasure and virtue they contain must have the same kind of goodness. Why should the virtue's having a distinct property of moral goodness also give it more ordinary goodness? If the former property is truly distinct, why should it bear on the latter? In any case, I do not think it best to treat moral goodness as a distinct property. The simpler view is that moral goodness is just intrinsic

⁸ See, for example, J. L. A. Garcia, "The Primacy of the Virtuous," *Philosophia* 20 (1990): 78–79; see also his "Goods and Evils," *Philosophy and Phenomenological Research* 47 (1987): 407–09.

goodness—the very same property—when had by certain objects, namely, attitudes evaluated in relation to their objects. It is just the goodness of higher-level rather than base-level goods.⁹ On this view, however, the claim that virtue alone is morally good supports no substantive view about virtue's comparative worth. To establish that worth, we must compare virtue directly with other values, such as pleasure, pain, and knowledge.

In my view, this comparison supports not Rashdall's view that virtue is a greater good, but the contrary view that virtue and vice are lesser values. This view is expressed in the following comparative principle about attitudes and their objects:

(CP) The degree of intrinsic goodness or evil of an attitude to x is always less than the degree of goodness or evil of x .

According to (CP), the intrinsic goodness of loving a good is always less than that of the good, as is the intrinsic evil of hating it. Desiring a pleasure is less good than the pleasure is, and being pained by the pleasure is less evil. Similarly, the evil of loving an evil is always less than that of the evil, as is the goodness of hating that evil. It is not, to repeat, that every virtuous or vicious attitude has less value than any instance of another good or evil, such as pleasure or pain. A desire for the immense pleasure of others, though less good than that pleasure, can be better than a mild pleasure of one's own. But every attitude does have less value than its specific object. Virtue and vice are nonlexicographically lesser values in the sense that they are always finitely less good or evil than the specific base-level states to which they are intentionally related.

On the simplest versions of (CP), every attitude has a maximum value that is the same fraction (for example, one half) of its object's value. No matter what the attitude's intensity, its goodness or evil cannot be more than one half its object's goodness or evil. In thinking about (CP), we need not commit ourselves to any such mathematically precise formulation, which in any case is unrealistic. (I doubt that any intrinsic values can be measured with this degree of precision.) But it will help to illustrate (CP) if we imagine a version on which attitudes have at most one half their objects' value.

There are several powerful arguments for (CP). The first, which derives from G. E. Moore, considers the combination of an intrinsic evil and a virtuous response to it—for example, one person's pain and another's compassion for that pain.¹⁰ If the value of a virtuous attitude were greater than that of its object, the combination of a pain and compassion for it would be on balance good, and better than if there were no pain and no compassion. But this, Moore claims, is unacceptable. The compassion is indeed good, and makes the situation better than if there were only

⁹ This simpler view is that of Ross and of William K. Frankena. See *The Right and the Good*, p. 155, and Frankena, *Ethics*, 2nd ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1973), p. 62.

¹⁰ *Principia Ethica*, p. 219.

pain and no compassion, but it cannot outweigh or justify the pain. Given a choice between creating a world containing only pain and compassion for it and creating nothing, it is surely best to create nothing. But if so, if a combination of pain and compassion for it is always on balance evil, then the positive value of virtuously hating an evil must be less than the negative value of the evil.

Or consider the combination of an evil attitude and a higher-level hatred of it—for example, a malicious desire to hurt another and shame at that desire. Since malice is evil by (LE), feeling pained by or ashamed of one's malice is good by (HE). This implication is attractive: that shame at a vicious desire is virtuous. But it would not be attractive to hold that feeling malice and shame at it is better than not feeling the malice at all. This conclusion would follow if the value of an attitude were greater than that of its object, but it does not follow given (CP).

A related argument concerns the combination of a good and vicious hatred of it—for example, person *A*'s pleasure and *B*'s envious wish that *A* were not experiencing that pleasure. *B*'s malicious envy is evil, but it surely cannot be so evil as to outweigh *A*'s pleasure or by itself make that pleasure undesirable. This conclusion is, again, avoided by (CP).

These initial arguments concern only some attitudes—namely, virtuous hatreds of evil and vicious hatreds of good. It is therefore possible to accept them while holding that other attitudes, such as virtuous loves of good, have more value than their objects. But it is surely simplest to hold that the relation between attitudes and their objects is uniform, so a conclusion like Moore's generalizes to all of (CP).

There are further arguments for (CP). One such argument concerns the value of very high-level attitudes. Given their recursive form, both (LG) and (HE) generate infinite hierarchies of intrinsically good attitudes. For example, (LG) values not only the love of pleasure but also the love of the love of pleasure, the love of the love of the love of pleasure, and so on to infinity. The clauses' recursive form is one of their attractive features. We have a better explanation of why *B*'s desire for *A*'s pleasure is good if we can say that love of any good is good, and not just the love of pleasure. It is also plausible in itself to hold that very high-level attitudes are good. But it would not be plausible to hold that these attitudes are very good or that we humans should spend much time trying to form them. (CP) avoids these implications. If the value of an attitude is always less than that of its object, say, by one half, the values of progressively higher-level attitudes get progressively smaller and diminish toward zero. Although *B*'s desire for *A*'s pleasure may have significant value, *C*'s desire for that desire of *B*'s has less value, and *D*'s desire for that desire of *C*'s even less. Very high-level attitudes still have some value, and it is a merit in, say, God that he can form them without detracting from his lower-level attitudes. But their value is sufficiently small that beings with our limited capacities should primarily work at developing lower-level virtues.

A third argument for (CP) invokes what I call the *proportionality view* about the division of virtuous love. Since we humans cannot love all good things with infinite intensity, the question naturally arises of how it is best to divide our love

between different good objects. The most appealing answer is: in proportion to the objects' degrees of goodness. If x is better than y , it is best to care more intensely about x but still to some degree about y . More specifically, if x is n times as good as y , it is best to care n times as intensely about x and $1/n$ times as intensely about y .¹¹ In this initial use, the proportionality view takes us from claims about the comparative values of objects to conclusions about the best divisions of love. But we can equally well use it in the opposite direction, saying that, if it is best to love x more than y , then x must be better than y . There are several instances of this argument for (CP).

Imagine that a teacher works to instill knowledge in a student from a genuine desire for the student's knowledge, and that as a result the student acquires knowledge. If you learn of these facts, which should you be more pleased by, the student's knowledge or the teacher's virtuous pursuit of knowledge? Surely, you should be more pleased by the student's knowledge; it is the point of the exercise. Or imagine that you can produce only one of these goods but not both. Right now an uncaring teacher is teaching using methods that do not work, so her student is not acquiring knowledge. You can either change the teacher's attitude while leaving her methods unchanged or change her methods but not her attitude. Surely, it is best to change the teacher's methods in this case so the student acquires knowledge. But then the student's knowledge must be better. Or consider an example involving evils. If a torturer is causing a victim intense pain and also taking sadistic pleasure in that pain, surely you should be most pained by the victim's pain. If you can either interfere with the torturer's pleasure while leaving his torture machine running or secretly disconnect the machine, you should disconnect the machine. If so, the evil of the torturer's pleasure in pain must be less than that of the pain.

Or consider attitudes to one's own virtue. What should the teacher in the example above have as her main motive for teaching, that she will thereby act virtuously or that through her teaching her student will gain knowledge? What should she be most pleased by afterward, her own good action or its effect? Many writers hold that there is a moral failing, called variously moral self-indulgence, narcissism, or priggishness, that consists in caring too much about one's own virtue. Bernard Williams says an agent is morally self-indulgent when "what the agent cares about is not so much other people, as himself caring about other people... [A] person may act from generosity or loyalty... and not attract the charge of moral self-indulgence, but that charge will be attracted if the suspicion is that his act is motivated by a concern for his own generosity or loyalty."¹² Williams

¹¹ For a fuller account of the proportionality view, see my "Self-Interest, Altruism, and Virtue," *Social Philosophy and Policy* 14 (1997): 286–307. Note that divisions of love that are slightly disproportionate are not evil. They are still good, just not ideally so.

¹² Bernard Williams, "Utilitarianism and Moral Self-Indulgence," in his *Moral Luck* (Cambridge: Cambridge University Press, 1981), p. 45.

does not hold that *any* concern for one's own virtue is a failing; self-indulgence consists only in caring "disproportionately" about one's virtue, or having it be "more important" in one's motivation than a concern for other people.¹³ He therefore allows that some concern for one's virtue, provided it is a lesser concern, is consistent with ideal attitudes. His view therefore fits perfectly with (CP). Because the teacher's virtuous action is intrinsically good, it is indeed appropriate for her to desire and be pleased by it. But if her action is less good than the knowledge it produces, what is best is for her to love her virtue less than she does the knowledge. If she cares more about her virtue, she divides her love disproportionately, which is precisely a moral failing.

Not all philosophers who condemn self-indulgence can do so on the basis of (CP). Ross,¹⁴ most notably, recognizes the "self-absorption" of an excessive concern with one's own character, but also holds that virtue has infinite value compared to other goods. The question, however, is whether a position like Ross's is consistent. It is surely plausible that given two goods, one greater than the other, it is best to care more about the greater than about the lesser. This is the core of what I have called the proportionality view about division. If we accept this view, however, and if virtue is even a finitely greater good, it follows that it is best to care more about one's virtue than about its effects. It is a commonplace of contemporary ethics that self-indulgence is a moral failing and that truly virtuous agents are not motivated primarily by thoughts of their own virtue. What is less often noticed is the tension between these claims and the still widely held view that virtue is the greatest among goods. In fact, the best way to explain why self-indulgence is a failing is to hold, with (CP), that virtue is a lesser good.

A final, more abstract argument for (CP) concerns virtue's status as a dependent intrinsic good. On the recursive account, virtue consists in certain attitudes to other goods and evils and therefore presupposes other goods and evils. It cannot be the only intrinsic value, but requires there to be base-level values to which it responds. It does not follow that virtue cannot be intrinsically good—the recursive account shows that it can—but a weaker conclusion does seem appropriate. If virtue is an attitude to other goods and evils, it seems fitting that it should be a lesser good, one whose value cannot exceed that of its object. Something secondary in the source of its value should likewise have secondary weight. There is no logical compulsion about this argument; it is perfectly consistent to hold that virtue is a response to other goods but, as that response, has infinitely more value. Nonetheless, it seems in a looser way appropriate that a good defined by reference to other values should be subordinate in its comparative standing.

I have given four types of argument for (CP): (1) about certain combinations of an object and an attitude, such as pain and compassion for it; (2) about very

¹³ Williams, "Utilitarianism and Moral Self-Indulgence," pp. 47, 45. See also Noah M. Lemos, "High-minded Egoism and the Problem of Priggishness," *Mind* 93 (1984): 550.

¹⁴ *Aristotle* (New York: Macmillan, 1949), p. 208.

high-level attitudes; (3) about the best divisions of love; and (4) about virtue's status as a dependent good. Together, these arguments strongly support (CP). (CP) holds that while any instance of virtue or vice may outweigh some instances of other goods or evils, it always has less value than the specific good or evil to which it is intentionally connected. This claim is one we can consider only given something like the recursive account of virtue and vice. If we treated the virtue concepts as primitive concepts rather than as connoting attitudes to other goods or evils, we could consider only lexicographic claims about virtue's general comparative worth. If we rejected those claims, we could not say anything in the abstract about virtue's rank among values. But if we take every instance of virtue and vice to be directed at another good or evil, we can compare its value to that of that specific good or evil. In this comparison, I have argued, virtue and vice are lesser intrinsic values.

III. A Lesser Good: Difficulties

There are, then, powerful reasons to accept the comparative principle (CP). But there are also difficulties about (CP) that a full discussion must address. Let me consider three such difficulties.

The first difficulty concerns Moore's argument about pain and compassion. (CP) implies that a combination of a pain and one person's compassion for it is always on balance evil. But what if two, twenty, or a hundred people feel compassion? Are the resulting combinations still on balance evil? I think our initial response is that they are. Our judgment may not be as confident here as in the case involving one person's compassion, but I think many of us will say that given a choice between creating pain plus a hundred people's compassion and creating nothing, it is best to create nothing. But it is hard to see how the recursive account even with (CP) can endorse this judgment. If each of the hundred people's compassion is good, then even if each is less good than the pain is evil, can the sum of their goodness not outweigh the pain? Can enough virtuous attitudes to pain not make for on-balance positive value? This conclusion would not follow if the value of extra people's compassion diminished the more people felt compassion, but there seems no reason to believe it does. That others are sympathizing with someone's pain makes it no less appropriate for me to do so and should therefore make it no less good. We may think the value of extra compassion diminishes if we consider compassion instrumentally, as helping to console the victim. (If ninety-nine people are already sympathizing with him, he will not take much more comfort from a one-hundredth.) But we do not think this way if we consider the compassion by itself. Then the value of each person's compassion seems the same, and this implies that enough people's compassion can outweigh any pain.

This difficulty weakens the Moorean argument for (CP). That argument claims that by accepting (CP), the recursive account can capture our intuitive judgments

about combinations of pain and compassion; if some such judgments are not captured, the support given (CP) is less. But it is not clear in what direction the difficulty points. It is certainly not a reason to hold that virtue is a greater good; that would only make the problem worse. If we cannot hold that the value of extra compassion diminishes, we may have to accept the conclusion about numbers, difficult though that may seem. We want to hold that virtue is intrinsically good; we also want to hold that in some cases, like that of the two worlds described above, virtue can outweigh base-level values, such as pleasure and pain. Given these assumptions, however, and adding that the value of extra people's compassion is constant, it seems unavoidable that some number of people's compassion can outweigh any pain. Although it is initially unsettling, the conclusion seems to be one we have to accept.¹⁵

The second difficulty concerns (CP)'s scope. As originally formulated, (CP) applies indiscriminately to all attitudes, saying they all have less value than their objects. This claim is compelling for attitudes whose objects are great goods or evils—for example, are intense pleasures or pains. But it is less so for attitudes whose objects are trivial goods or evils. Imagine that *B* goes out of his way to provide from benevolence a small pleasure for *A*, say, by doing *A* a small favor. May in this case *B*'s benevolence not have more value than the pleasure it produces? (Of gift giving we say, "It is the thought that counts.") Or imagine that *B* takes special trouble to cause *A* a small annoyance or pain. May his petty malice not be worse than its effect? In general, may attitudes to trivial objects not have greater value than their objects?

If we were persuaded by these examples, we would have to restrict (CP) so it applies only to attitudes whose objects have significant value. More generally, we could revise (CP) so the ratio between the value of an attitude and that of its object is not constant, as in the simple versions discussed above, but changes continuously with the object's value. When the value of the object is high, the value of the attitude is a small fraction of it. As the object's value declines, the fraction gets larger; and when the object is trivial, the attitude's value is greater. Instead of a constant ratio, such as one half, between the value of an attitude and that of its object, there are different ratios for objects of different values.

Unfortunately, this revised (CP) is incompatible with the proportionality view about the best divisions of love. That view, recall, holds that if *x* is *n* times as good as *y*, it is best to love *x* *n* times as intensely as *y*. The simplest justifications of this view assume that if *x* is *n* times as good as *y*, loving *x* with a given intensity is always *n* times as good as loving *y* with that intensity.¹⁶ This last assumption can

¹⁵ We may also have to accept a parallel conclusion about time, namely, that a sufficiently long-lasting compassion can outweigh any pain. This conclusion may be mitigated to some extent if we hold that the degree of compassion appropriate to a pain, and therefore the value of that compassion, diminish as the pain recedes into the past. But it is not clear that this condition will always prevent long-lasting compassion for a short-lived pain from being more good than the pain itself is evil.

¹⁶ See the discussions of the "asymptotic" and "optimality" views in my "Self-Interest, Altruism, and Virtue."

be maintained if the value of an attitude is always a constant fraction of its object's value, as on the simple versions of (CP). But it cannot be maintained given the revised (CP). The revised principle therefore requires abandoning the proportionality view in favor of one on which the best division of love between x and y is not strictly proportioned to their values but gives slightly more love to the lesser good y than its degree of goodness would suggest. Some may be willing to accept this departure from strict proportionality and may therefore adopt the revised (CP). But I find the original proportionality view highly appealing and therefore prefer to retain the simpler (CP). This requires accepting what some may find counterintuitive, that even attitudes to trivial goods and evils have less value than their objects. (In gift giving, it is the gift that counts.) So each of the available options is in some respect questionable. But whichever option we choose, the recursive account will retain the core claim of (CP): that for objects above a threshold of goodness or evil, the value of any attitude to them is less than their own.

The most important difficulty for (CP) concerns the value on balance of certain vices of loving evil, such as sadistic pleasure in another's intense pain. Imagine that a torturer takes pleasure in the pain he causes his victim. The recursive account says, through its recursion clause (LE), that his pleasure is intrinsically evil as an instance of loving evil, or of vice. But it also says, through its base clause (BG), that his pleasure is intrinsically good as an instance of pleasure, or as having the pleasantness that makes all pleasure sensations as such desirable. The torturer's pleasure is therefore intrinsically evil as pleasure in pain but intrinsically good as pleasure. So is it on balance intrinsically good or intrinsically evil? We surely want it, intuitively, always to be on balance evil, so its presence always makes a situation worse. In fact, its ability to make such pleasure intrinsically evil is a central attraction of the recursive account as against purely subjective views, such as hedonism. But whether sadistic pleasure is on balance evil depends on the comparative values of its qualities of vice and pleasantness. Only if the pleasure's evil as vice always outweighs its goodness as pleasure will it always be on balance evil. How can this be, given (CP)? Imagine that (CP) makes the maximum value of an attitude always one half that of its object, and that equal units of intensity of pleasure and pain have equal value; then, if the torturer takes two units of pleasure in his victim's ten units of pain, his pleasure may be on balance evil. It can have more than two units of disvalue as vice and therefore make the overall situation worse. But if he takes six units of pleasure in his victim's ten units of pain, his pleasure has more goodness as pleasure than its maximum five units of vice, and makes the situation better. If it is sufficiently intense, a sadistic pleasure can be, counterintuitively, on balance good.¹⁷

¹⁷ This difficulty is nicely illustrated in Frankena's *Ethics*. He claims both that malicious pleasure is always on balance evil and that every pleasure has some value as pleasure (pp. 89–90). But he does not explain why, as this combination of claims assumes, the moral evil of malicious pleasure must always outweigh its hedonic goodness as pleasure.

This difficulty would not arise given the lexicographic view, which makes virtue and vice infinitely more valuable than pleasure or pain. Ross cites this fact in arguing for the view: “If the goodness of pleasure were commensurable with the goodness or badness of moral disposition, it would be possible that such a pleasure [that is, a pleasure of cruelty] if sufficiently intense should be good on the whole. But in fact its intensity is a measure of its badness.”¹⁸ As we have seen, however, there are compelling reasons to reject the lexicographic view. Once virtue and vice have just finite value compared to their objects, let alone lesser value, it seems that some instances of pleasure in intense pain can be on balance good.

Can the recursive account avoid the implication that sadistic pleasure can be on balance good? One way is to deny that pleasure as such is intrinsically good, or belongs among the base-level values identified by (BG). This antihedonist view can allow that pleasures with good objects are good; they are still virtues as valued by (LG). But it denies that pleasure considered independently of any object, or in itself, is good.

This antihedonist view certainly avoids the claim that pleasure in evil can be on balance good; if doing so were a necessity, we might have to adopt it. But the view has many other counterintuitive implications. For example, it denies that nonintentional pleasures, such as those of suntanning or eating ice cream, are good. Although not completely unimaginable, this is an extreme step. Surely, most of us believe that simple bodily pleasures have some intrinsic worth. A less sweeping antihedonism denies only that intentional pleasures are good as pleasures: unstructured bodily pleasures have value, and only pleasures that something is the case do not. But this view still denies, implausibly, that pleasures with neutral intentional objects are good. Consider a sports fan’s pleasure, which may be very intense, that his local team has won a championship. If his feeling involves the same quality of pleasantness as the pleasure of suntanning, should it not likewise be good? Nor do only pleasures with neutral objects seem good as pleasures; those with good objects do as well. Imagine that *A* lives in a world with many very good objects in which he takes appropriate pleasures, whereas *B* lives in an evil world by which he is appropriately pained. Do we not think that *A*’s state, though no more virtuous than *B*’s, is on balance better because his virtue can take the form of pleasure?

It may seem that the solution to these difficulties is obvious: deny only that pleasures with evil intentional objects are good as pleasures. Nonintentional pleasures are good as pleasures, as are pleasures with neutral or good objects. Only pleasures in evil objects, such as sadistic pleasures, lack value as pleasures. Their orientation to evil has two effects, making these pleasures evil as attitudes and canceling their goodness as pleasures. But attractive though it may seem, this moralized antihedonism would require major changes in the structure of the recursive account and even then cannot be captured in its entirety.

¹⁸ *The Right and the Good*, p. 151.

In the original recursive account, the base clauses (BG) and (BE) are prior to and independent of the recursion clauses (LG)-(HE), identifying their non-moral goods and evils without reference to any subsequent claims about good and evil attitudes. The base clauses are also independent of each other, with neither restricted by the other's claims. But these structural features exclude the moralized antihedonism just proposed. To say that sadistic pleasure has no value as pleasure is to make a claim about base-level values. To say it has no value because it is directed at evil is to ground that claim in something recursive; and in the original recursive account, no base-level claim can depend on any recursive claim.

Because of this, capturing the proposed antihedonism would require changing the account's structure, first by abandoning the independence of (BG) from (BE). A revised account starts with the original (BE), which says that all pain, false belief, and failure are evil. It then follows (BE) with a version of (BG) that is restricted by (BE), one saying that pleasure is good *except when* it is directed at the evils identified by (BE)—that is, except when it is directed at pain, false belief, or failure. This revised (BG) implies, as desired, that sadistic pleasure has no value as pleasure; when the recursion clause (LE) is added, sadistic pleasure will always be on balance evil. The revised (BG) also allows that pleasures with no, neutral, or good objects are good. So it has at least initially the right kind of implications. But it does not capture all the proposed antihedonism, since it does not imply that pleasures with higher-level evil objects lack value as pleasures. Imagine that while *B* takes sadistic pleasure in *A*'s pain, *C* takes a higher-level pleasure in *B*'s sadistic pleasure, enjoying the fact of *B*'s sadism. Surely, if *B*'s pleasure lacks value as pleasure, *C*'s should as well; it, too, is directed at evil. But a (BG) restricted only by (BE) cannot endorse this claim, since it denies value only to pleasures in base-level evils. If it is to yield the claim, the recursive account must be revised further to abandon the priority of all base-level over all recursive claims. More specifically, it must interpose the recursion clause (LE) about hating evils between (BE) and (BG). Then (BG) can be doubly restricted, saying that pleasure is good except when it is directed at the evils identified by either (BE) or (LE). This second revision makes *C*'s pleasure in *B*'s sadistic pleasure lack value as pleasure, as desired. But it still does not capture all the proposed antihedonism. Imagine that *B* hates a good of *A*'s—for example, is enviously pained by *A*'s pleasure—and that *C* takes a higher-level pleasure in *B*'s envious pain, delighting in it because it is envious. Here, *C* loves an evil attitude of *B*'s, as in the previous example, but this attitude is now one of hating good rather than loving evil. Since the envious pain that *C* loves is evil by recursion clause (HG), *C*'s pleasure in it should not on our current proposal have value as pleasure. But this last claim cannot be captured by any changes in the recursive account. The clause that makes *B*'s envious pain evil, (HG), concerns the hatred of goods, and unlike (LE), which presupposes only base-level evils, (HG) requires a prior identification of base-level goods. But this characterization is supposed to be given by the very base clause (BG) that we are

now trying to restrict! Capturing the desired claim about pleasure in envious pain therefore involves a circularity: (BG) is to be restricted by reference to a recursion clause (HG) that in turn requires a completed (BG) before its own claims have substance. Changing the structure of the recursive account can capture some of the view that pleasures with evil objects lack value as pleasures, but it cannot without circularity capture all of it.

The original structure of the recursive account, with independent base clauses prior to any recursion clauses, is attractive in its simplicity. Because of this, any change that complicates the account's structure is in that respect unattractive. It is especially so if the change does not result in a fully motivated or principled view. Surely, what is plausible here is only the fully general view that all pleasures with evil objects lack value as pleasures; any more partial view is unacceptably ad hoc. So, if the recursive account cannot capture all the proposed moralized antihedonism, because it cannot say that pleasures whose objects are made evil by (HG) lack value as pleasures, that antihedonism should be abandoned.

Where does this leave us? In my view, there is only one principled way for the recursive account to hold that all pleasure in pain is on balance evil, and that is to deny that pleasure as such is a base-level good. This is not an unimaginable move, but it has the extreme implication that innocent pleasures, such as those of sunbathing or supporting a sports team, have no intrinsic value. If we cannot accept this implication, we must grant that even pleasures in pain have value as pleasures and can therefore sometimes be on balance good. But this may not in the end be an unacceptable claim.

First, the claim parallels one that is positively attractive. The recursive account holds through base clause (BE) that compassionate pain—that is, pain at another's pain—is always evil as pain. Given (CP), it therefore also holds that compassionate pain can sometimes be on balance evil. But this is not an unattractive implication. On the contrary, it fits our practice of sometimes not revealing our hurts to our friends, to spare them the pain of sympathizing with us. We may recognize that our friends' compassion would be good as virtuous, but we seem sometimes to view its evil as pain as more significant. But if in this case the moral goodness of an attitude can be outweighed by its hedonic evil as pain, why cannot in a parallel way the moral evil of pleasure in pain be sometimes outweighed by its hedonic goodness?

Second, there are cases where it seems right to hold that pleasure in pain is on balance good. Imagine that *B* takes pleasure in some minor mishap of *A*'s, such as *A*'s slipping on a banana peel, or laughs at a slightly pointed joke at *A*'s expense. (A great many jokes, including most political jokes, turn in a slightly malicious way on some person's failings.) On the recursive account, there is something morally objectionable about *B*'s pleasure; it would be better if he got the same enjoyment from a neutral or, better, a good object. But what if he cannot do so? What if the only options given *B*'s character are his getting pleasure from slight evils for *A* and his getting no pleasure at all? Is it then best if *B* gets no pleasure? This seems an

excessively prudish view.¹⁹ It is, to repeat, unfortunate that *B* can get pleasure only in this way; it would be better if he had nonmalicious sources of enjoyment. But to say that his pleasure is on balance evil is to condemn, all things considered, what is for many people one of their readiest sources of enjoyment. The issue here is not the value of pleasures in great evils, such as a torturer's pleasure in his victim's intense pain. It is only that of pleasures in small evils, ones the love of which (CP) makes an even smaller evil. It does not seem wrong to hold that these pleasures' small evil as vices is outweighed by their greater goodness as pleasures. On the contrary, it seems right, on all but a prudish view.

Pleasures in small evils, then, can be on balance good. But what about pleasures in great evils, such as the torturer's sadistic pleasure in his victim's intense pain? Is it not still counterintuitive to hold that they can be on balance good? It is, but the recursive account can agree that these pleasures are always on balance evil if it revises its treatment of pleasure as a base-level good, and in particular denies that equal units of intensity of pleasure and pain always have equal value.

The base clauses (BG) and (BE) imply that pleasures and pains are, respectively, better and worse when they are more intense, but they do not say how much they are better and worse. The simplest view says the relationships here are linear: the value of an additional unit of intensity of pleasure or pain is always the same, so pleasures and pains that are twice as intense are always twice as good or evil. This linear view is attractive for the intensity of pains, and I shall assume at least initially that the account applies it to pains. But the view is considerably less attractive for pleasures. Imagine that *A* has a choice between the certainty of a pleasure of ten units of intensity and a gamble with a .51 probability of giving him a pleasure of twenty units of intensity and a .49 probability of giving him no pleasure. The linear view says he should prefer the gamble, but I think most of us would say he should not. Or imagine that *C* can give either *A* a pleasure of twenty-one units of intensity or each of *A* and *B* a pleasure of ten units. The linear view says *C* should give *A* the twenty-one units but, again, many would disagree. To capture the judgments these examples suggest, the recursive account can abandon a linear treatment of pleasure and hold instead that the value of an extra unit of intensity in a pleasure gets smaller as the pleasure's intensity increases, diminishing asymptotically toward zero. On this asymptotic view, the second ten units of intensity in a pleasure have less value than the first, so *A* should not take his gamble and *C* should give the equal pleasures to *A* and *B*. Assuming these are attractive implications, we have independent reasons to adopt the asymptotic view; and doing so can avoid the implication that pleasures in great evils can be on balance good.

If the recursive account retains a linear view of pain, and holds that the maximum value of an attitude is always the same fraction of its object's value, it holds that the moral evil of loving a pain that is twice as intense is always twice as

¹⁹ For a discussion of jokes that seems to assume this prudish view, see Ronald de Sousa, *The Rationality of Emotion* (Cambridge, Mass.: MIT Press, 1990), ch. 11.

great. On any view, this implies that the intensity of pleasure required to outweigh this moral evil is also greater. But given the asymptotic view of pleasure, the value of increases in the intensity of a pleasure gets progressively less. This implies, initially, that many fewer cases of pleasure in pain are on balance good than given a linear view of pleasure. More importantly, it implies that some pleasures in pain never can be on balance good. Since the asymptotic view makes the value of increases in the intensity of pleasure diminish toward zero, it places an upper limit on the value of any pleasure, whatever its intensity. This implies that, if the moral evil of a pleasure in pain is above that limit, its moral evil cannot be outweighed by its hedonic goodness. Imagine that the upper limit on the value of a pleasure is ten units, and that the value of an attitude of intensity i is always one half that of its object. Then, if B takes pleasure of intensity i in A 's thirty units of pain, the value on balance of his pleasure must be negative. Its moral evil as an attitude to its object is fifteen units, which is greater than the maximum ten units of goodness it can have as a pleasure. So his vicious pleasure makes the situation worse. By combining a linear treatment of pain with an asymptotic treatment of pleasure, the recursive account can allow that some pleasures in mild pains are on balance good. This, I have argued, is a plausible implication. But it can also hold that pleasures in more intense pains, ones above a certain threshold, are always on balance evil. By treating pleasure and pain asymmetrically, it can consistently hold that pleasure in all forms is a base-level good, that attitudes always have less value than their objects, and that pleasures in great pains are never on balance good.

The more specific implications of this approach depend on exactly how fast the value of extra units of pleasure diminishes, or exactly what the upper limit on a pleasure's value is. If this limit is low, comparatively few cases of pleasure in pain are on balance good, but intense innocent pleasures have only limited value, compared to other goods. If the limit is higher, innocent pleasures have more value, but more pleasures in pain are good. It may be that no precise version of the approach is attractive in every respect: either it gives what seems too little value to innocent pleasures, such as those of suntanning, or it allows too many vicious pleasures to be good. This dilemma may be relieved to some degree if the recursive account extends its asymmetrical treatment of pleasure and pain by adopting a nonlinear view of pain contrary to the one adopted for pleasure. On this view, the value of an additional unit of intensity in a pain is not constant but gets greater as the pain's intensity increases, so pains that are twice as intense are more than twice as evil.²⁰ This second nonlinear view makes intense pain more evil, which given (CP) makes sadistic pleasure in it also more evil as vice; and this allows the upper limit on the value of a pleasure as pleasure to be higher while pleasures in intense pain are still always on balance evil. This doubly nonlinear approach may not entirely resolve the dilemma about good versus evil pleasures. Nonetheless, either it or the simpler

²⁰ This view is defended in Jamie Mayerfeld, "The Moral Asymmetry of Happiness and Suffering," *Southern Journal of Philosophy* 34 (1996): 317–38.

asymmetrical view described first seems to offer the best available response to the difficulty about the value on balance of sadistic pleasure, allowing the recursive account to retain its original structure, affirm (CP), and still deny that the worst pleasures in evil can be on balance good. In fact, this is an additional reason to embrace the asymmetry: it offers the only way of reconciling, within a recursive framework, the overwhelmingly plausible (CP) and the intuitive view that sadistic pleasures are always on balance evil.

IV. Conclusion

Let me summarize the main points of this chapter. I have considered the comparative intrinsic values of virtue and vice, understood as perfectionist values, and have done so within a recursive account of virtue and vice that allows certain non-lexicographic claims about them to be formulated. I have argued that virtue and vice are lesser intrinsic values in the sense that their intrinsic goodness or evil is always less than that of their specific intentional object. Although instances of virtue and vice may outweigh some instances of other goods and evils, they always have less value than the specific one to which they are intentionally connected. Finally, I have considered certain difficulties for this comparative view, arguing that they do not provide good reasons to abandon it. That virtue and vice are lesser values is not without contestable implications, but it is the best available view on this topic.

Postscript

A longer version of this essay appeared as chapter 5 of my book *Virtue, Vice, and Value*.

Two Kinds of Organic Unity

In *Principia Ethica*, G. E. Moore defends a *principle of organic unities* that he states as follows: “The value of a whole must not be assumed to be the same as the sum of the values of its parts.”¹ If two states are brought together so a certain relation holds between them, the resulting whole may have either more or less intrinsic value than the states would have if they existed alone. This principle of organic unities is attractive and implicit in a great many views about intrinsic value. In this paper I will distinguish two interpretations of the principle, a holistic interpretation that Moore himself employed and a conditionality interpretation that he rejected. Despite differing on some apparently important issues, these interpretations can always reach the same overall conclusions about value—that is, find the same total value in a given state of affairs. For practical purposes, therefore, the choice between them is not important. But there are nonetheless differences between the interpretations that make sometimes one and sometimes the other more appropriate for expressing a given evaluative view. And in at least two pairs of cases, involving deserved pain and undeserved pleasure, and compassionate pain and malicious pleasure, the choice between them is morally important. Let me start by distinguishing the two interpretations.

II. Holism vs. Conditionality

The generic principle of organic unities shared by the interpretations can be stated more explicitly as follows:

For helpful comments on an earlier version of this chapter I am grateful to the members of the Ethical Theory Research Group at the University of Calgary, the participants at the Conference on Intrinsic Value at UNC/Greensboro, and, especially, my conference commentator Peter Vallentyne.

¹ G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), p. 28.

The intrinsic value in a whole composed of two or more parts standing in certain relations need not equal the sum of the intrinsic values those parts would have if they existed alone, or apart from those relations.

To illustrate this principle, consider a case central to Moore, the admiring contemplation of beauty. In chapter 3 of *Principia Ethica* Moore argues that the existence of beauty apart from any awareness of it has intrinsic value, but in chapter 6 he allows that beauty on its own at best has little and may have no intrinsic value.² And in the later work *Ethics* he implicitly denies that beauty on its own has value.³ But Moore always holds that the admiring contemplation of beauty has value. He also holds that when someone admiringly contemplates what he believes is real beauty, that beauty's actually existing and causing his contemplation makes the overall situation considerably better, or better by more than can be attributed to any value in the beauty itself. Let us assume Moore's later view that the existence of beauty on its own has zero value, and that the value of admiringly contemplating a given beauty that one believes to be real is m . Moore holds that the existence and causal role of that beauty makes for a whole whose value is greater than m (i.e., is $m + n$). This illustrates the generic principle, since the sum of the values the parts of this whole would have outside it is $m + 0$.

A view structurally similar to Moore's has been defended by William K. Frankena, Derek Parfit, and on what I think is the best interpretation of his doctrine of higher pleasures, John Stuart Mill. According to this view, perfectionist states such as knowledge, achievement, and even aesthetic appreciation have no value if they are not accompanied by some positive attitude toward them, such as desire for or pleasure in them, for their own sakes. These positive attitudes may have value apart from the states themselves. Thus, desire for or pleasure in knowledge may have value apart from any knowledge. (It is familiar that a desire for knowledge can exist in the absence of knowledge. So can pleasure in knowledge if a person takes pleasure in what he falsely believes is knowledge; consider Aristotle's taking pleasure in what he falsely believed was his explanatory knowledge of biology.) But if one of the positive attitudes is accompanied by the real existence of the perfectionist state that is its object, the resulting value is greater than the attitude would have on its own. The value of knowledge on its own is zero, the value of pleasure in knowledge on its own is m , but the value of knowledge plus pleasure in knowledge is $m + n$.⁴

² G. E. Moore, *Principia Ethica*, pp. 189, 202, 203.

³ G. E. Moore, *Ethics* (Oxford: Oxford University Press, 1965), p. 107.

⁴ William K. Frankena, *Ethics*, 2nd ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1973), pp. 89–92; Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 501–02; and John Stuart Mill, *Utilitarianism*, ed. Oskar Piest (Indianapolis: Bobbs-Merrill, 1957), pp. 12–16. A similar view has been discussed in political philosophy under the heading the “endorsement thesis”; see Ronald Dworkin, “Foundations of Liberal Equality,” *The Tanner Lectures on Human Values*, vol. 11 (Salt Lake City: University of Utah Press, 1990), pp. 1–119; and Will Kymlicka, *Contemporary Moral Philosophy* (Oxford: Clarendon Press, 1990), pp. 203–04. Some presentations of this view conflate two distinct questions.

Moore's presentation of the principle of organic unities assumes a specific interpretation of it, one that is forced on him by his definition of intrinsic value. According to this definition, which I will call a strict definition, a state's intrinsic value can depend only on its intrinsic properties, that is, properties it has independently of any relations to other states. As Moore puts it, when value is intrinsic, "the question whether a thing possesses it, and in what degree it possesses it, depends solely on the intrinsic nature of the thing in question."⁵ This strict definition implies what Noah M. Lemos calls the "thesis of universality": if a state's intrinsic value can depend only on its intrinsic properties, it follows that that value must be the same wherever the state appears and, in particular, whatever wholes it enters into.⁶ Moore explicitly accepts this implication: "The part of a valuable whole retains exactly the same value when it is, as when it is not, a part of that whole."⁷ But his accepting the universality thesis forces Moore to a *holistic* interpretation of the principle of organic unities. If the parts of a whole retain the same value in the whole as they had outside it, the whole's value can differ from the sum of its parts only if it contains an additional value that results from their being combined in that specific way. This is exactly Moore's view. He holds that if a person admiringly contemplates a beauty that exists and causes his contemplation, then the contemplation has the value it would have on its own (namely m), and the beauty has the value it would have on its own (namely zero), but there is an additional value n in the whole composed of contemplation-plus-beauty-plus-causal-relation, for an overall value of $m + n$. This additional value n is intrinsic in Moore's strict sense. Since it is an intrinsic property of the contemplation-plus-beauty-plus-causal-

One is whether a pleasure intentionally directed at a perfectionist state such as knowledge is better than a pleasure with no such object, for example a bodily pleasure such as that of eating ice cream. The other question is whether the combination of a pleasure directed at a perfectionist state and the real existence of that state is better than the pleasure alone. Only the second of these questions involves the principle of organic unities (see the end of section II in this chapter). The conflation is evident, for example, in Parfit's suggestion that there is a symmetrical dependence between pleasure and perfectionist states, so that just as knowledge without pleasure lacks value, so pleasure without an object like knowledge lacks value. But the pleasure that is necessary for knowledge to have value is not just any pleasure; it is pleasure in knowledge and not just, say, an accompanying pleasure of eating ice cream. And if we ask whether pleasure in knowledge has value in the absence of knowledge, the answer seems to be "yes." Aristotle's pleasure in what he falsely believed was his knowledge of biology seems every bit as good as pleasure in real knowledge of the same value. The pleasure that lacks value on Parfit's view is pleasure without an object like knowledge, but that pleasure is not the condition for knowledge's having value. And the pleasure that is the condition for that value does not lack value on its own.

⁵ G. E. Moore, "The Conception of Intrinsic Value," in his *Philosophical Studies* (London: Routledge & Kegan Paul, 1922), p. 260. See also Moore's use of a "method of isolation," whereby we test for intrinsic value by asking whether a world containing only a given state and no other to which it could be related is good (*Principia Ethica*, pp. 93, 95, 187–88; *Ethics*, pp. 24, 68). The strict definition is also adopted in Christine M. Korsgaard, "Two Distinctions in Goodness," *Philosophical Review* 92 (1983): 169–95; Roderick M. Chisholm, *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986), pp. 52–53; and Noah M. Lemos, *Intrinsic Value: Concept and Warrant* (Cambridge: Cambridge University Press, 1994), pp. 8–11.

⁶ Lemos, *Intrinsic Value*, p. 11.

⁷ Moore, *Principia Ethica*, p. 30.

relation that it involves those elements related in that way, the additional value ascribed to it depends on its internal nature. Moore's approach is made most explicit in his distinction between the intrinsic value of a whole "as a whole" and its intrinsic value "on the whole."⁸ A whole's intrinsic value as a whole is its intrinsic value just as a combination of parts and independently of any values in those parts. Its intrinsic value on the whole is its intrinsic value on balance or all things considered—that is, the value that results from adding its intrinsic value as a whole to any values in its parts. In the aesthetic case, the whole composed of the contemplation, the beauty, and the causal relation between them has more intrinsic value on the whole than the sum $m + o$ of the values of its parts because it has an additional intrinsic value n as a whole that must be added to them. A similar analysis can be given of Frankena's view: the whole composed of pleasure in knowledge and knowledge has more value on the whole than the pleasure and the knowledge because it has an additional value as a whole, or as pleasure-in-knowledge-plus-knowledge.

This holistic interpretation of the principle of organic unities is in one sense a weak one. Though it denies that the value of a whole always equals the sum of the values its parts would have outside it, it does not deny that the value of the whole is always a sum of which those values are components. The alternative *conditionality* interpretation does deny this. It holds, against Moore, that the intrinsic value of a state can change when it enters into a larger whole, so its value or degree of value is altered by its relations to other states. On this interpretation the value of a whole need not equal the sum of the values the parts would have outside it, not because there is an additional value in the whole as a whole, but because the parts change value when they enter it.

The conditionality interpretation allows a different formulation of the views discussed in holistic terms above. Applied to Moore's later view about aesthetic value, it says that although beauty on its own has no value, the same beauty when it is the object of admiring contemplation does have value. When someone admiringly contemplates beauty that exists and causes his contemplation, the overall value in the situation can be, as on the holistic interpretation, $m + n$. But the additional value n is now located not in the contemplation-plus-beauty-plus-causal-relation but in the beauty itself. Similarly, in Frankena's view the overall intrinsic value of pleasure in knowledge plus knowledge can be the same as on the holistic formulation, but the additional value now is a conditional value in the knowledge.

Because it rejects the universality thesis, the conditionality interpretation must also reject the strict definition of intrinsic value that implies that thesis. It must hold, more loosely, that a state's intrinsic value is that portion of the overall value of the world that is attributable to or located in that particular state. This value will often be determined by the state's intrinsic properties, but not always. Sometimes a

⁸ Moore, *Principia Ethica*, pp. 214f.

state's intrinsic value in the looser sense can depend on its relations to other states, or its place in larger wholes.

It may be objected that what I have called a looser definition of intrinsic value is not really a definition of intrinsic value at all. In an influential article, Christine M. Korsgaard claims that when a state's value does not derive solely from its intrinsic properties it is really extrinsic value, since it is value the state gets "from some other source."⁹ But Korsgaard's claim is exaggerated for the particular views we are considering. In the conditionality version of Moore's later view, not just anything becomes good when it is the object of admiring contemplation; only beauty does and ugliness does not. In Frankena's view knowledge becomes good when it is the object of pleasure but ignorance does not. And beauty and knowledge have this status because of their intrinsic natures; their intrinsic properties make them such that when they stand in certain relations they are good. Though the source of their value is not purely internal, it is also not purely external.

If we were still impressed by Korsgaard's claim, we might coin the term "midtrinsic" value for value that has this mixed basis, but I do not see what point other than pedantry this would serve. When it actually exists, conditional value plays the same role as uncontentiously intrinsic value. It contributes directly to the overall value of the world; it is something we should care about and pursue for its own sake because of its value. I therefore see no reason to rule out the looser definition of intrinsic value or to deny that what the conditionality interpretation ascribes conditionally to states such as beauty and knowledge is intrinsic value.

I said above that the conditionality interpretation is in one sense stronger than the holistic interpretation, since it does not hold that the value of a whole is always a sum of which the values the parts would have outside it are components. But it is in another sense weaker. If the values of states can change when they enter a whole, there is no need to deny that the value of the whole always equals the sum of the values the parts have when inside the whole, or in that particular context. The conditionality interpretation can, if it wants, deny this. It can hold that sometimes when the values of parts change there is also an additional value in the whole as a whole, and we will later discuss a view with this mixed structure.¹⁰ But unlike the holistic view, the conditionality view can equally well hold that the value of a whole always equals the sum of the (perhaps changed) values its parts have inside it.

We have, then, two interpretations of the principle of organic unities. One starts from the strict definition of intrinsic value, accepts the universality thesis,

⁹ Korsgaard, "Two Distinctions in Goodness," p. 170. Korsgaard's claim may be true for the view that states are good whenever they are the object of desire, or of rational desire. But that view is very different from the views I am expressing using the principle of organic unities.

¹⁰ It is logically possible for a mixed view to deny the generic principle of organic unities. Thus, it is logically possible to hold that whenever the value of a part changes when it enters a whole, there is a compensating value in the whole as a whole that restores the overall value to the sum of the values the parts would have alone. But surely this is just a logical possibility. Surely anyone who makes conditionality claims will want them to have some effect on the overall values of wholes.

and finds additional values in wholes as wholes to the values their parts would have on their own. The other accepts the looser definition of intrinsic value, denies universality, and holds that the values of parts can change when they enter wholes. These interpretations seem to disagree on some fundamental issues in value theory, such as the nature of intrinsic value and the thesis of universality. These look like issues it is vital to settle, and they are certainly issues philosophers have debated. But in the contexts we have considered the disagreements cancel each other out and have no practical significance. Both interpretations can agree that although an attitude to an object has value m and that object on its own has value zero, the combination of attitude and object has value $m + n$. More generally, the interpretations can always agree on the overall intrinsic value in a given whole. Whatever conclusion the holistic interpretation reaches by ascribing a value to the whole as a whole, the conditionality interpretation can reach by changing the values of the parts.

II. Three Differences

If the two interpretations can always agree in their conclusions, the choice between them is not important for practical purposes such as calculating the overall intrinsic value in a situation or the world. Nonetheless, there are at least three differences between them that make sometimes one and sometimes the other more appropriate. Some views that presuppose the generic principle of organic unities seem better stated in holistic terms, while others seem better stated conditionally.

The first difference concerns the location of the additional value in an organic unity, which the holistic interpretation places in the whole as a whole but the conditionality interpretation places in a part. This difference matters especially if we think of intrinsic goods as states it is appropriate to care about—for example, to desire or take pleasure in, for their own sakes—since then the two interpretations have different implications for where our attitudes should be directed.

In Moore's later and Frankena's views this difference may not matter much if we consider the attitudes of an observer. If person *A* admiringly contemplates beauty that exists and causes his contemplation, it may not matter whether an observer *B* is directed to take pleasure in the contemplation-plus-beauty or in the beauty itself. But it does seem to matter if we consider *A* himself. On the holistic interpretation, what *A* admires, namely the beauty, is not itself ever intrinsically good. What is good is only a whole containing the beauty, never the beauty itself. This means that the attitude of *A*'s that Moore's view makes necessary for additional value seems to involve a kind of mistake. It is directed not at a good, but only at something that though part of a good is itself intrinsically neutral. This mistake is not present to the same degree on the conditionality interpretation. Then what *A* admires is, at least when he admires it, intrinsically good and a fit object of admiration. If this makes the evaluative importance

Moore ascribes to *A*'s admiration more intelligible, it favors the conditionality over the holistic interpretation of Moore's later view.¹¹

There are other views where the location of value seems important. Several philosophers hold that if a person pursues a good goal, the successful achievement of that goal, even after his death, makes his pursuit intrinsically better than if his efforts had ended in failure. For example, if he works for the preservation of Venice, and if partly through his efforts Venice is preserved after his death, his activities are better than if Venice had been destroyed.¹² This view is usually presented in a context that is partly egoistic. It is assumed that a person has reason to care if not exclusively then at least more about goods in his own life. But this assumption goes better with a conditionality than with a holistic interpretation. On the former interpretation, the extra value that results from the successful preservation of Venice is located in the person's activities and therefore in his life. It is conditional on a causal relation between his life and events outside it, but is nonetheless a value in him. On the holistic interpretation, by contrast, the additional value is located not in his life but in a whole combining activities in his life and events after his death. And why should that whole be a special concern of his? The holistic view could be combined with a restatement of partial egoism telling people to care more about goods either in their lives or in wholes of which events in their lives are parts. But this is not the intuitively most natural statement of egoism, which normally mandates extra concern for one's own life. Nor does holism give the most natural statement of the view about posthumous achievement, which is that realization of a goal after one's death makes one's own activity better. In fact, this argument for conditionality seems to generalize to all instances of achievement and knowledge. Both these goods involve a relation between a mental state and a state in the world that somehow matches it: in achievement, an intention in a person's mind is realized in the world; in knowledge, a belief in his mind accurately represents the world. But both achievement and knowledge are normally seen as goods of a person or in his life. This is possible only if intentions and beliefs, which are states internal to the person, acquire additional value when they stand in certain relations to states outside him. It is not possible if the value in achievement or knowledge is ascribed to a whole containing a mental state, an external state, and the relation between them.¹³

¹¹ Note that there is not this reason to favor the conditionality interpretation of Moore's earlier view, on which beauty on its own does have intrinsic value. On this earlier view *A*'s admiration is always of something good in itself, and its appropriateness as an attitude can be equally well explained by a holistic interpretation.

¹² Parfit, *Reasons and Persons*, p. 151; I discuss this view on what is implicitly a conditionality interpretation in *Perfectionism* (New York: Oxford University Press, 1993), pp. 110–11.

¹³ A different account of these issues is proposed by Shelly Kagan. He holds that when the value of a person's life is affected by a relation between himself and a state of the world such as the preservation of Venice, that relational fact and perhaps also the state of the world are internal to or parts of his life, even though they are not internal to him (Shelly Kagan, "Me and My Life," *Proceedings of the Aristotelian Society* 94 [1993/94]: 318–23). But if we can use the looser definition of intrinsic value, as Kagan himself

A second difference between the interpretations is suggested by a claim of Korsgaard's. She defends a conditionality view of the value of happiness, saying happiness is good only when it is morally deserved, or is the happiness of a person with a morally good will or with what I will call virtue. Korsgaard recognizes that a holistic or Moorean formulation of this type of view is also possible, but she rejects it as "perverse" on the ground that it "veil[s] or obscure[s] the internal relations within the organic unity in virtue of which the organic unity has its value."¹⁴ As stated, Korsgaard's claim is simply false. The holistic interpretation values the very same relations—for example, that of beauty's causing admiring contemplation—as the conditionality interpretation. It sees those relations as holding internally, between the parts of a whole, rather than externally, between separate states. But it values and emphasizes the same relations between the same relata.

Nonetheless, Korsgaard's claim does point to a relevant difference. The holistic interpretation treats the parts of a whole symmetrically, ascribing value to an entity, namely the whole as a whole, of which they are both equally and in the same way parts. The conditionality interpretation, by contrast, treats the parts asymmetrically. It assigns the additional value in a whole to one of the whole's parts, but only on condition that the other part is present. If that other part has value on its own, the parts relate to intrinsic value differently; one contributes to value unconditionally, the other only when a certain condition is satisfied.

I think this difference is a further reason for preferring the conditionality interpretation of Moore's later and Frankena's views. In these views the admiring contemplation of beauty and pleasure in knowledge are good independently of any connection to real beauty or knowledge, but the beauty and knowledge are not good independently. This asymmetry is reflected in a conditionality formulation, which makes one part of the whole depend for its contribution to value entirely on the other. The difference in the parts' relation to value is mirrored in the way value is ascribed to them. This argument for conditionality does not apply to Moore's earlier view, where beauty on its own does have some intrinsic value. Here the parts of the whole contribute symmetrically to the whole, since both have value on their own and both must be present for additional value. This symmetry makes, if anything, the holism Moore adopted most appropriate. And there are other views whose symmetry makes a holistic interpretation appropriate.

Consider an account of vices of disproportion such as extreme selfishness, where a person cares much more about his own good than about the equal or greater goods of other people.¹⁵ Extreme selfishness can involve two attitudes that

allows, I see no reason to make this claim. We can equate a person's life with a sequence of states or events within his body and mind and still say that the value of that life is affected by relations to external states even though neither those states nor the facts relating him to them are internal to his life.

¹⁴ Korsgaard, "Two Distinctions in Goodness," pp. 192–93.

¹⁵ See Thomas Hurka, "Self-Interest, Altruism, and Virtue," *Social Philosophy and Policy* 14 (1997): 295–98.

on their own are good. Intense concern for one's own good is good, as to a lesser degree is mild concern for others' good, at least if it is above some threshold intensity. But if mild forms of malice and callousness, where one desires or is indifferent to another's minor evil, are vicious and evil, surely extreme selfishness is also evil. This view can be captured given the generic principle of organic unities, since then there can be an additional evil when two good but disproportionate attitudes are brought together. But a conditionality interpretation of the view seems to involve a certain arbitrariness. We can hold a selfish person's intense concern for his own good constant and say that, given that concern, his merely mild concern for others' good is evil. Or we can hold his mild concern for others' good constant and say that his intense self-concern is evil. But there seems no non-arbitrary way to choose between these formulations. What is more, they both miss how the distinctive evil in extreme selfishness is not located in either attitude individually but is a matter of their combination as a combination. We describe this evil best in a holistic interpretation. Then there is an additional evil in any combination of attitudes whose intensities are out of proportion to the values of their objects as a combination, an evil that is added to any values the attitudes have on their own. If the disproportion is small, as in mild selfishness, the evil is also small and may be outweighed by the goodness in the attitudes. As the disproportion increases, however, the evil becomes greater and eventually makes the person's combination of attitudes evil not just as a whole, but also on the whole or on balance.

A similar argument applies to what Roderick M. Chisholm calls the "*bonum progressionis*," where a sequence containing the same total goodness in its individual parts is better if it progresses from the less good to the better than if it declines from the best to the worst.¹⁶ A conditionality interpretation of this view again faces an arbitrary choice: do we say the earlier parts of an improving sequence have more value because they will be followed by something better, or that the later parts have more value because they were preceded by something less good? And either choice misses how what gives the sequence its extra value is its shape as a whole, the very thing that is the explicit focus of a holistic interpretation.

The view Korsgaard discusses, about deserved and undeserved happiness, is harder to assess by these lights. At first glance, desert as she conceives it seems to involve the same asymmetry as Moore's later and Frankena's views. Virtue is good regardless of any connection to happiness, whereas happiness is good only given virtue. But there is a subtle difference. In Moore's and Frankena's views the value of the attitude is not only unconditional, but in a way complete: if pleasure in knowledge is accompanied by knowledge that makes the situation better, but only in the way that any additional good would make the situation better. But in the desert view the value of virtue is not complete. Virtue seems to call for happiness as a complement, or as something that goes especially with it. Virtue plus happiness

¹⁶ Chisholm, *Brentano and Intrinsic Value*, pp. 70–71.

is not just virtue plus some other good, but virtue plus its own fitting reward. There is therefore a kind of symmetry in the desert view, albeit one where the mutual dependences of the two components are different. This may explain why some philosophers formulate their desert views not conditionally, as Korsgaard does, but holistically. Both Moore and W. D. Ross hold that the apportionment of happiness to virtue is a good present in a whole composed of a certain degree of virtue and a certain degree of happiness as a whole, and to be added to any further goodness in the virtue and happiness on their own.¹⁷

These first two differences between the holistic and conditionality interpretations are not, I concede, very significant. They concern at best small advantages of one over another formulation of essentially the same evaluative view. The third difference is more substantial. It arises when the additional value in a whole is opposite to the value of one of the parts—for example, is positive when the part's value is negative.

This kind of case is illustrated by negative or retributive desert, where a person's vicious character or action makes him deserve to suffer pain. A holistic interpretation says that if a vicious person suffers deserved pain there are three intrinsic values present: the initial evil of the vice, the further evil of the pain, and the good of retribution in the whole composed of these two evils as a whole. If the punishment is to be on balance worth inflicting, the goodness of the retribution must be greater than the evil of the pain, but it must not be so great that the combination of the vice, pain, and punishment is better than if there had been no vice at all.¹⁸ The important point, however, is that on the holistic interpretation there are at the point of punishment two intrinsic values being created, the good of retribution and the evil of pain, and that our attitudes to what is happening should therefore be mixed. We should be pleased that desert is being done but pained that this involves pain, with the former attitude stronger but the latter still present. This is an attractive implication. In an insightful discussion Robert Nozick draws a number of contrasts between the desire for retribution and the desire for revenge. One concerns their accompanying emotional tone: revenge involves pleasure in another's suffering, retribution does not.¹⁹ But I think we can go further and say that retribution involves or should involve a distinctively somber or subdued emotional tone, one suffused with regret. The holistic interpretation can explain why. At the point of punishment we should be pleased that justice is being done but pained at the infliction of pain, with the latter emotion qualifying and limiting the first.

This explanation is not available given a conditionality interpretation of desert. This interpretation, which has been elaborated more fully by Fred Feldman, holds that desert is not an additional value to pleasure and pain but a factor that "adjusts" their values, so pleasure and pain have more value when deserved and less when

¹⁷ Moore, *Principia Ethica*, p. 215; W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 58, 72, 138–40.

¹⁸ Moore, *Principia Ethica*, p. 215.

¹⁹ Robert Nozick, *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981), p. 367.

undeserved. In some cases a hedonic state's value can even be "transvalued," so it has the opposite value to the one it would have on its own. Retribution illustrates this phenomenon: when a person suffers pain as deserved punishment, his pain is not evil but has been transvalued to be good.²⁰ Feldman's conditionality interpretation can reach the same overall conclusions about retribution as the holistic interpretation, but it claims that at the point of punishment only one value is being created, namely the transvalued good of the pain. And this makes it difficult to explain the distinctively somber tone that ought to accompany retribution. If all that is being created is something good, should the appropriate emotion not be pure and simple pleasure?

The case of deserved pain, then, seems to favor a holistic view about desert. But some will think the opposite conclusion is supported by the contrary case of undeserved pleasure, where a vicious person who should suffer pain enjoys delight. A holistic interpretation implies that there are two values created with the pleasure: the goodness of the pleasure and the evil of the undesert. The evil may outweigh the good, so the pleasure is on balance undesirable, but the good is still present. But I think some philosophers will deny that the pleasure of a vicious person is in any way good, and they will especially deny that our response to such pleasure should be mixed, including some pleasure in the pleasure. On the contrary, they will say that if a thoroughly vicious person enjoys pleasure our response should be only displeasure at the undesert. The issue here is not so much the interpretation of claims about desert as such. To deny that undeserved pleasure is good as pleasure, one need not adopt a conditionality view such as Feldman's, on which the pleasure itself is evil. One can retain, perhaps for reasons of symmetry, a holistic view such as Moore's and Ross's on which what is evil is a whole composed of vice and pleasure as a whole. What matters is the treatment of the pleasure. Applying a conditionality view just to that, one can say that undeserved pleasure lacks value because undeserved, so the only value created with such pleasure is the holistic evil of the undesert. One's view will then be mixed, holding that there can simultaneously be an additional value in a whole as a whole and a changed value in a part. But the view will include as one element the conditionality claim that pleasure loses value when combined with moral vice.

A similar tension arises for the related cases of pleasures and pains that are themselves virtuous or vicious. Consider compassionate pain, where one person is pained

²⁰ Fred Feldman, "Adjusting Utility for Justice: A Consequentialist Reply to the Objection from Justice," in his *Utilitarianism, Hedonism, and Desert* (Cambridge: Cambridge University Press, 1997), p. 167. Feldman's desert view differs from Korsgaard's in holding that pleasure in the absence of desert has, not zero, but some positive value, and also in denying that virtue on its own has value. Like any conditionality view, Feldman's presupposes the looser definition of intrinsic value on which a state's value can be affected by its relations to other states. But in his "On the Intrinsic Value of Pleasures" (*ibid.*, pp. 125–47), Feldman criticizes a "Sidgwickian" view of the intrinsic value of pleasure for being inconsistent with the strict definition. His own theory of desert seems to appeal to the very understanding of intrinsic value he denies to views that ground the value of pleasures in a relation to desires.

by another's pain. This pain seems good in so far as it involves compassion, but evil in so far as it involves pain. Our response to it should therefore be mixed: pleasure that the person is compassionate, but pain that he feels pain. We may even think in some cases that the evil outweighs the goodness; consider how we sometimes do not tell our friends our troubles to spare them the pain of sympathizing with us. But the contrary case of malicious or sadistic pleasure, where a person takes pleasure in another's pain, will again strike some as different. They will deny that a torturer's pleasure in his victim's intense pain is in any way good, or that we should feel, alongside pain at the torturer's sadism, any pleasure in his pleasure. The situation parallels the one just discussed about desert. Both deserved and virtuous pain seem to remain evil as pain, but some will deny that undeserved and vicious pleasure remain good as pleasure. Whereas the first item in each pair calls for a treatment that keeps the evil of the pain constant, some will say the second requires a conditionality view on which the goodness of pleasure is cancelled by its relation to undesert or vice.

I want to discuss these two pairs of cases more fully, since they are where the choice between a conditionality view and some rival is most important. But first a qualification is needed. Though deserved pain and undeserved pleasure clearly are organic unities, compassionate pain and malicious pleasure are not, on at least the most natural understanding of the parts and wholes relevant to such unities. On this understanding these items are particulars—that is, particular objects or states of affairs involving the existence of particular objects. And malicious pleasure, to take just that example, is not a whole combining distinct particulars. It is a single state, a pleasure that is intentionally directed at pain and would not be the specific pleasure it is were it not directed at pain. (Remember that a pleasure in pain can exist without any pain, as when a person takes pleasure in what he falsely believes is real pain.) That malicious pleasure is an organic unity is claimed by Chisholm, Lemos, and others, but only on the basis of what I think is a strained understanding of parts and wholes.²¹ The more natural view is that malicious pleasure is a single state with two properties relevant to its intrinsic value. It is a pleasure, which tends to make it good, and a pleasure in pain, which tends to make it evil. If its value on balance is just the sum of the values these properties always give states that have them, there is no departure from simple summation and nothing like an organic unities view. Nonetheless, one can take something like a conditionality view of

²¹ These philosophers hold that *B* is a proper part of *A* whenever *A*'s existence conceptually implies *B*'s but *B*'s existence does not conceptually imply *A*'s (Chisholm, *Brentano and Intrinsic Value*, p. 73; Lemos, *Intrinsic Value*, pp. 33–34; Michael J. Zimmerman, "Virtual Intrinsic Value and the Principle of Organic Unities," *Philosophy and Phenomenological Research* 59 [1999]: 655). This makes a person's feeling pleasure a part of his feeling pleasure in pain, they claim, since he cannot feel pleasure in pain without feeling pleasure but can feel pleasure without feeling pleasure in pain. But this analysis requires understanding parts and wholes as universals. It is not true that the person could feel the particular pleasure he feels without feeling pleasure in pain, since that is what his particular pleasure is. What is true is only that he could feel some pleasure or other without feeling pleasure in pain. But this treatment of parts and wholes as universals seems to me strained, as more generally does the claim that its being a pleasure is part of a pleasure in pain rather than an aspect or property of it.

such pleasures. One can hold that the value its being a pleasure gives a state depends on what other properties the state has, and in particular is zero when among those properties is its being a pleasure in pain. This view extends to the properties of a single state something very like the treatment given different states by the conditionality interpretation of the principle of organic unities. The view is one that, as I have said, some will find intuitively appealing on the same general ground as they find a conditionality view of undeserved pleasure appealing. So we can discuss the two views together, and that is what I will do. I will ask whether we should hold that undeserved and malicious pleasure lose value because of their connection to vice or whether they remain good as pleasures.²²

III. Undeserved and Malicious Pleasure

These pleasures are sometimes discussed as test cases for the general thesis of universality,²³ but I hope my earlier discussion has shown that this is a mistake. In other contexts universality can be either accepted or rejected with little practical effect, and some views seem best stated without it. One can therefore reject universality elsewhere but accept what it implies for undeserved and malicious pleasure, because of specific facts about these pleasures. I will now argue that this should be done. The best accounts of these pleasures treat them as, though perhaps evil on balance, nonetheless good as pleasures.

My discussion will assume that we cannot take a conditionality view of the related cases of deserved and compassionate pain, but must treat these as, though good as deserved or compassionate, at the same time evil as pain. Only so can we explain the mixed attitudes that are appropriate to these pains. This suggests a first argument against a conditionality view of undeserved and malicious pleasure,

²² Korsgaard ascribes the conditionality view of undeserved pleasure to Immanuel Kant and endorses it herself; she would presumably endorse the parallel view about malicious pleasure. In the most extensive recent discussion of these views Lemos attributes both views to Ross (Lemos, *Intrinsic Value*, pp. 41–42), but this seems to me on several grounds a mistake. First, Ross's explicit claims that undeserved and malicious pleasure are evil (*The Right and the Good*, pp. 136–38) are open to three interpretations: (i) such pleasures are purely evil with no element of good; (ii) such pleasures are evil as undeserved or malicious, though perhaps also good as pleasures; (iii) such pleasures are evil on balance, with their evil as undeserved or malicious outweighing their goodness as pleasures. I see nothing in Ross's text that clearly favors (i) over (ii) and especially (iii), but only (i) expresses a conditionality view. Second, Ross accepts Moore's strict definition of intrinsic goodness (*ibid.*, pp. 114–21), which implies the universality thesis and logically prevents pleasure from changing value when undeserved. Third, a related argument of Ross's explicitly assumes a denial of the conditionality view. Ross holds that virtue and vice have infinite value compared to pleasure and pain, and one of his main arguments for this view concerns malicious pleasure. It claims that unless virtue has infinite comparative value, malicious pleasure could if sufficiently intense be on balance good, which Ross thinks is impossible (*ibid.*, p. 151). This argument presupposes that malicious pleasure remains good as pleasure, with that goodness always needing to be outweighed by a greater evil of malice.

²³ See, e.g., Lemos, *Intrinsic Value*, pp. 40–47.

based on symmetry. If in the first case in each pair the negative hedonic value of pain remains alongside the goodness of retribution or compassion, should the positive hedonic value of pleasure not likewise remain alongside the evil of undesert or malice? I do not think too much weight can be placed on this argument, since there is no guarantee that the best theory of value will be pleasingly symmetrical. But whatever reason there is for favoring simpler over more complex theories does support extending a nonconditionality treatment from the first to the second case in each of our two pairs.

A more substantial argument against the conditionality view of undeserved pleasure has recently been proposed by several writers, including Lemos, Michael Slote, and Robert Audi. It claims that we cannot explain what is evil about undeserved pleasure unless we assume that such pleasure is good, or a benefit to its possessor. As Lemos puts it, "What is it that makes the wicked man's being happy...so offensive? I suggest that it is offensive precisely because we think that the wicked man has a *good* that he deserves not to have...the judgement that the prosperity of the wicked is not good...presupposes the judgement that his being pleased *is* good."²⁴ I think this argument is best expressed given a holistic interpretation of claims about desert. It then says that a whole composed only of vice and pleasure in the same person is not one that is evil as a whole. Only if we take the whole to combine vice and something good for the vicious person is it as a whole objectionable.

Appealing as this argument seems, it has two important limitations. First, it does not apply at all to the related case of malicious pleasure, since it is not true that we can explain why such pleasure is evil only if we hold that it is in one respect good. Malicious pleasure involves a positive attitude to something evil, and is evil because it involves such an attitude. But there are other positive attitudes to evil, such as a desire for or pursuit of something evil, that are evil on the same basis. And to hold that a desire for something evil, such as a desire for another's pain, is evil we do not have to hold that desire as such is good, nor is it even plausible to hold that desire as such is good. So the parallel argument to Lemos's has no force. Yet given the close similarity between undeserved and malicious pleasure, one would expect a successful argument against a conditionality view of one to extend to the other.

Second, the argument's central claim about undesert does not extend to wholes combining vice and goods other than pleasure. Imagine that a vicious person has knowledge, which many consider an intrinsic good, or at least has more knowledge than a virtuous person. Does the combination of vice and the good of knowledge strike us as evil as a whole, or as involving an objectionable kind of undesert? I do not think it does. Innocence as traditionally understood combines virtue and

²⁴ Lemos, *Intrinsic Value*, pp. 43–44. See also Michael Slote, "The Virtue in Self-interest," *Social Philosophy and Policy* 14 (1997): 273–74; and Robert Audi, "Intrinsic Value and Moral Obligation," *Southern Journal of Philosophy* 35 (1997): 140.

ignorance, while a common picture of vice allows it to involve a sly or clever ability to achieve its evil ends. Reflecting on these pictures, we may think the innocent person lacks a good the vicious person has, but I do not think we find any additional evil of unfairness in this fact. We do not care about combinations of vice and knowledge in the way we do about combinations of vice and pleasure. This point is even clearer if we consider wholes in which the second good is virtue. We need to remember that views about desert do not consider only wholes combining simultaneous goods and evils. Retribution can be served when a person who acted viciously in the past is punished later, and the pleasure that rewards virtue can likewise come later. So imagine that one person is now virtuous while another is vicious, and consider two possible futures. In one the two people retain their present characters but in the other they switch, so the virtuous person becomes vicious and the vicious person virtuous. If wholes combining vice and any future good were evil as wholes, the first future would be better than the second, since a vicious person who later became virtuous would be getting a good he did not deserve. But I do not think we have any tendency to make this judgment or to have any preference between the futures. Whatever is true of combinations of vice and pleasure, it is not true that every combination of vice and something good is evil.

This point can be expressed in terms of the common distinction between what is deserved and its “desert basis,” or that feature of a person in virtue of which he deserves it.²⁵ Normally a given desert basis makes only some items deserved and not others. Virtue calls for pleasure or happiness but not money; economic contribution or effort calls for money but not happiness. It would be an interesting project to try to explain the connection between different desert bases and what specifically they make deserved, but the important point is that there is a connection. If a person is virtuous, he deserves pleasure but not all possible goods; if he is vicious, what makes for an evil of undesert is again not all goods.²⁶

In an article on organic unities it would be rash to draw too strong a conclusion from this point. Lemos and other proponents of the argument we are considering could hold that just as pleasure on its own does not when combined with vice make for an evil whole, so goodness on its own does not make for an evil whole. But when pleasure and goodness are combined, so a person has pleasure that is also good, that combination when added to vice makes for an evil of undesert. There is a prior organic unity needed for the organic unity of undesert, one combining pleasure and goodness into a whole that alone is such that when added to vice it makes for a larger evil. But I do think this second limitation weakens the

²⁵ Joel Feinberg, “Justice and Personal Desert,” in his *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, N.J.: Princeton University Press, 1970), pp. 55–94.

²⁶ It may be that different goods are called for by different forms of virtue. For example, if a person is dedicated and principled in his pursuit of knowledge, perhaps what he deserves is knowledge more than pleasure. But the general point remains. For any form of virtue only some goods are appropriate rewards, and for any form of vice only some goods make for an evil of undesert.

argument considerably. Perhaps if pleasure is needed for an evil of undesert it does not follow that the pleasure does all the work in creating that evil, but the claim that it does not is a very fine-grained one on which it is hard to place the whole weight of an argument against a conditionality view of undeserved pleasure. When we add that the argument does not extend to the related case of malicious pleasure, I think we do best to look for another ground for holding such pleasures to be good as pleasures.

I believe there is such a ground, which can be stated first for the case of malicious pleasure and then extended to undeserved pleasure. It claims that the best explanation of why malicious pleasure is evil as malicious cannot be given, or at least cannot plausibly be given, unless we assume that all pleasure is as pleasure good.²⁷

Earlier I said that malicious pleasure is evil because it involves a positive attitude toward something evil, as do desire for and pursuit of something evil. This explanation is best given within a general recursive theory of the intrinsic values of attitudes shared by many writers on intrinsic value, from Brentano, Moore, and Ross to Nozick, Chisholm, and Lemos.²⁸ Let me sketch this theory briefly.

Like any recursive theory this one needs base clauses, which in this case affirm that certain states other than attitudes are intrinsically good and evil. It does not matter what the exact content of these clauses is, but let us assume, at least initially, that the theory begins with the following base clauses about intrinsic goods and evils:

(BG) Pleasure and knowledge are intrinsically good.

(BE) Pain and false belief are intrinsically evil.

The theory then adds four recursion clauses about the intrinsic values of attitudes to intrinsic goods and evils. Two clauses affirm the intrinsic goodness of morally appropriate attitudes:

(LG) If x is intrinsically good, loving x (desiring, pursuing, or taking pleasure in x) for itself is intrinsically good.

(HE) If x is intrinsically evil, hating x (desiring or pursuing x 's non-existence or being pained by x 's existence) is intrinsically good.

Two further clauses affirm the intrinsic evil of inappropriate attitudes:

(LE) If x is intrinsically evil, loving x for itself is intrinsically evil.

(HG) If x is intrinsically good, hating x for itself is intrinsically evil.

²⁷ I also give this argument in Thomas Hurka, "How Great a Good Is Virtue?" *The Journal of Philosophy* 95 (1998): 181–203.

²⁸ See Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, trans. Roderick M. Chisholm and Elizabeth Schneewind (London: Routledge & Kegan Paul, 1969), pp. 22–23; Moore, *Principia Ethica*, pp. 204, 208–209, 211, 217; Ross, *The Right and the Good*, pp. 160–161; Nozick, *Philosophical Explanations*, pp. 429–33; Chisholm, *Brentano and Intrinsic Value*, pp. 62–67; Lemos, *Intrinsic Value*, pp. 34–37, 73–77; and Thomas Hurka "Virtue as Loving the Good," *Social Philosophy and Policy* 9 (1992): 149–68.

The part of this theory that explains why malicious pleasure is evil is clause (LE), about the evil of loving evils. And I think the explanation is deepened by being embedded in a more general recursive theory that makes many attitudes, not just to evils but also to goods, and with higher-level as well as base-level objects, intrinsically good and evil.

Why does this explanation require all pleasure to be good as pleasure? The reason is that the recursive theory has to start with base clauses affirming that certain states other than attitudes are intrinsically good and evil. Only given these base clauses can we know which specific attitudes the later recursion clauses identify as also good and evil. My particular base clause (BG) holds that pleasure is a base-level good, which implies that all pleasure, including malicious pleasure, is good as pleasure. Now, some may take this implication precisely as a reason to reject my (BG), but what can they offer in its place? They can deny that pleasure is ever a base-level good,²⁹ but I do not think many will be attracted to this move. It implies that even morally innocent pleasures, such as those of eating ice cream or watching a sunset, lack value as pleasures, and I do not think many will accept that. Instead, those who want to deny that malicious pleasure is good as pleasure will try to do so on a more restricted, conditionality, basis. They will hold that pleasure is good as pleasure except when it is pleasure in an evil such as pain. Pleasure with a good, a neutral, or no intentional object is good as pleasure, but pleasure with an evil object, pleasure that is made evil by recursion clause (LE), is not.

But this conditionality view is inconsistent with the basic structure of the recursive theory, which requires base-level values to be identified before higher-level ones and independently of them. The proposed view attempts to specify the base-level good of pleasure by reference to higher-level considerations about appropriate and inappropriate attitudes, and this inevitably leads to a circularity. What must come first in the theory is made to depend on what must come later, and this cannot be done in a principled or thoroughgoing way.

In my version of the recursive theory the base clauses (BG) and (BE) are stated both before the recursion clauses and independently of each other. We could capture part of the proposed conditionality view by changing this order, stating (BE) about evils before (BG) about goods and restricting the latter by the former, so (BG) now says that pleasure is good except when it is pleasure in one of the specific evils identified by (BE). This reordering implies, as desired, that pleasure in pain is not good as pleasure. But it does not capture all the proposed conditionality view, since it does not imply that pleasures in higher-level evils, those made evil by the recursion clauses, lack value as pleasures. Imagine that while *B* takes malicious pleasure in *A*'s pain, a third person *C* takes pleasure in *B*'s malicious pleasure, delighting in it because it is malicious or as a pleasure in pain. Surely if *B*'s malicious pleasure lacks value as pleasure, *C*'s pleasure should too; it too has an evil object.

²⁹ Zimmerman seems to propose this in "Virtual Intrinsic Value and the Principle of Organic Unities," p. 13.

But a (BG) restricted only by (BE) cannot capture this judgment. If we wanted to capture it, we could reorder the theory further. We could interpose recursion clause (LE) about loving evils between (BE) and (BG) and make the latter doubly restricted, so it now says that pleasure is good except when it is pleasure in one of the evils identified by either (BE) or (LE). This further reordering implies, as desired, that *C*'s pleasure in *B*'s malicious pleasure as malicious is not good as pleasure. But it still does not capture the whole conditionality view, and here we come to the ineliminable circularity. Imagine that *A* enjoys pleasure and that *B* is enviously pained by *A*'s pleasure, wishing *A* were not enjoying that pleasure and feeling unhappy that he is. Imagine further that *C* takes pleasure in *B*'s envious pain, delighting in it as envious or as a pain in pleasure. Here *C* loves an evil attitude of *B*'s, as in the previous example, but this attitude is now one of hating good rather than loving evil. Since the envious hatred that *C* takes pleasure in is evil, *C*'s pleasure should not on the conditionality view be good as pleasure. But this last judgment cannot be captured without circularity. The clause that makes *B*'s envious pain evil is (HG) about hating goods, and unlike (LE), which presupposes only a prior identification of base-level evils, (HG) presupposes a prior identification of base-level goods. But this identification is supposed to be given by the very (BG) the conditionality view is now trying to restrict! We cannot hold that *B*'s envious pain at *A*'s pleasure is evil unless we have already determined that *A*'s pleasure is good. And this means that, unlike (LE), (HG) cannot be stated before (BG) and used to restrict it. We cannot say that pleasure is good as pleasure except when it is pleasure in a hatred of a lower-level good, because we would have to know first whether pleasure is one of those lower-level goods. We would be trying to restrict (BG) by an (HG) that requires (BG) to be complete before its own claims have substance. And that is circular.

It is impossible, then, to include a generalized conditionality view of pleasures in evil in the recursive theory without circularity. We can capture an arbitrarily restricted part of the view, but not one with a thoroughgoing rationale. But exactly how much does this prove? It would not prove much if there were some other explanation of the evil of malicious pleasure available aside from the recursive one, but I do not see what this other explanation could be. Consider the Kantian view that pleasure in another's pain is evil because it fails to show respect for that other person's autonomy or rational nature. This view has to explain why disrespect is shown in particular by pleasure in pain, and I do not see how it can explain that without appealing to the fact that the pain is an evil for its subject. Surely that is the central moral fact about malicious pleasure—that it involves pleasure in something evil—and this means that an adequate explanation of its evil must invoke the very recursive structure that rules out generalized conditionality.

For the same reason, this argument against a conditionality view of malicious pleasure extends to undeserved pleasure. Undeserved pleasure is pleasure combined with vice, which means that an explanation of its evil presupposes an account of what vice is. But surely any account of vice must include as instances of

vice those loves of evil such as pleasure in pain, pleasure in malice, and pleasure in envy, and surely it can do so only given something like the recursive theory. An explanation of the evil of undeserved pleasure therefore presupposes the same recursive structure as one of the evil of malicious pleasure and in the same way excludes a generalized conditionality view.

The conclusion of this argument may strike some as counterintuitive. They may prefer intuitively a view on which undeserved and malicious pleasure lack value as pleasure and therefore call only for pain at them rather than a mixed emotional response. But their resistance to the conclusion might be eased if it were always the case that such pleasures, though good as pleasures, were evil on balance, or weighing their good and evil qualities against each other. Then the appropriate attitude to them would be, though not purely, still predominantly one of pain. This claim is not difficult to establish for the case of undeserved pleasure. Just as the good of retribution can be given enough weight as against pain to make punishment always on balance justified, so the evil of undesert can be made to always outweigh the relevant pleasure. There is a greater difficulty about justifying the parallel claim about malicious pleasure, given a plausible general thesis about the comparative values of attitudes and their objects. But given an independently attractive view about the comparative values of pleasure and pain, we can hold that at least all pleasures in great evils, for example, in intense rather than in mild pains, are evil on balance.³⁰

There are, then, arguments for preferring a holistic treatment of claims about desert and a parallel nonconditionality view about malicious pleasure. But, to return to the themes of the earlier part of this chapter, this is not because of general theses about intrinsic value or some general superiority of holism over conditionality. In all cases the two interpretations of the principle of organic unities can yield the same substantive conclusions, and though in some cases there are reasons to prefer one over the other, these reasons vary from case to case. The issue between holism and conditionality always concerns the best treatment of a particular evaluative view that involves organic unities, not the grander theses about value that have been the focus of so much writing in this area.

³⁰ See Hurka, "How Great a Good Is Virtue?"

Asymmetries in Value

Values typically come in pairs. Most obviously, there are the pairs of an intrinsic good and its contrasting intrinsic evil, such as pleasure and pain, virtue and vice, and desert and undesert, or getting what one deserves and getting its opposite. But in more complex cases there can be contrasting pairs with the same value. Thus, virtue has the positive form of benevolent pleasure in another's pleasure and the negative form of compassionate pain for his pain, while desert has the positive form of happiness for the virtuous and the negative form of pain for the vicious.

Of each pair we can ask how its elements relate to each other, and the simplest answer is that they do so symmetrically—so that, for example, a pleasure of a given intensity is exactly as good as a pain of that intensity is evil, or benevolence exactly as great a virtue as compassion. But there is no necessity for this. Values can equally well be asymmetrically related, and in several ways. In this paper I ask, of a series of pairs of value, when asymmetries between their elements are plausible, what bases these asymmetries have, and whether there are any patterns among them. I start with the simplest case, that of pleasure and pain.

Utilitarians typically treat these values as symmetrical, so a given quantity of pleasure exactly cancels the disvalue of an equal quantity of pain; this is Jeremy Bentham's view and is also suggested by Henry Sidgwick.¹ But G. E. Moore disagrees. In *Principia Ethica* he says that while pleasure has "at most some slight intrinsic value," pain is "a great evil," adding that "[t]he study of Ethics would, no doubt, be far more simple, . . . if . . . pain were an evil of exactly the same magnitude as pleasure is a good; but we have no reason whatever to believe that the Universe is such that ethical truths must display this kind of symmetry."² A similar view has recently been defended by Jamie Mayerfeld. One of his two main claims about

¹ Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907), p. 413.

² G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), pp. 212, 222.

happiness and suffering is that it is more important to prevent suffering than to promote happiness, because “suffering is more bad than happiness is good.”³ And the view he shares with Moore is intuitively appealing. If one can either relieve one person’s intense pain or give a slightly more intense pleasure to another person, many will say it is right to relieve the first person’s pain. Or imagine first a world containing only intense mindless pleasure like that of the deltas and epsilons in *Brave New World*. This may be a good world, and better than if there were nothing, but it is surely not very good. Now imagine a world containing only intense physical pain. This is a very bad world, and vastly worse than nothing.⁴

The Moore-Mayerfeld view needs to be distinguished from two others. One holds that pleasure is nothing real but only the absence of pain; if we think it has positive qualities, we are only being fooled by the transition from a greater to a lesser pain. This view has been asserted by Plato, Epicurus, and Schopenhauer, but I will assume that it is false and that the pleasure of eating an oyster or of passionate love involves more than just feeling no pain. The second view holds that pleasure, though real, has no value at all. This is the view of negative utilitarianism, on which our only moral duty is to prevent pain. But negative utilitarianism implies that if we could either bring about a world in which billions of people are ecstatically happy but one person suffers a brief toothache or bring about nothing, we should bring about nothing. The Moore-Mayerfeld view avoids this absurd implication by giving happiness some positive value, so enough of it will outweigh minor pain. But exactly what does the view say?

The symmetry view says that a pleasure of a given intensity is always exactly as good as a pain of that intensity is evil. Its simplest version is represented in Figure 7.1, whose vertical axis measures the value of a pleasure or pain just as a pleasure or pain and whose horizontal axis measures its intensity, and which displays the relation between the two as a single straight line. (A pleasure or pain can also have nonhedonic value, for example if it is virtuous and on that basis good or undeserved and on that basis evil. Both Figure 7.1 and our current discussion abstract from these possibilities and consider the value of hedonic states only as hedonic

³ Jamie Mayerfeld, *Suffering and Moral Responsibility* (New York: Oxford University Press, 1999), p. 136.

⁴ The Moore-Mayerfeld view may seem more problematic when applied to choices within a single life. If a person prefers to trade off pain for himself for a slightly greater amount of pleasure for himself, are we to say he is wrong? And will we interfere to prevent him from making that choice? First, even if his preference is wrong, respect for his autonomy may make paternalistic interference with the preference wrong. Second, Mayerfeld suggests that the pleasure-pain asymmetry may be less strong within lives than across them, so a quantity of pleasure that would not outweigh a given pain for another can outweigh the same pain for oneself (*Suffering and Moral Responsibility*, pp. 150–51). Finally, and subject to this last qualification, some self-regarding tradeoffs of pain for pleasure may indeed be wrong. (The symmetry view certainly thinks so.) In this chapter I assume that value does not depend entirely on people’s desires, so someone who prefers mindless pleasure to knowledge and achievement has a mistaken preference. If so, someone who trades off pain for only slightly more intense pleasure may likewise be mistaken.

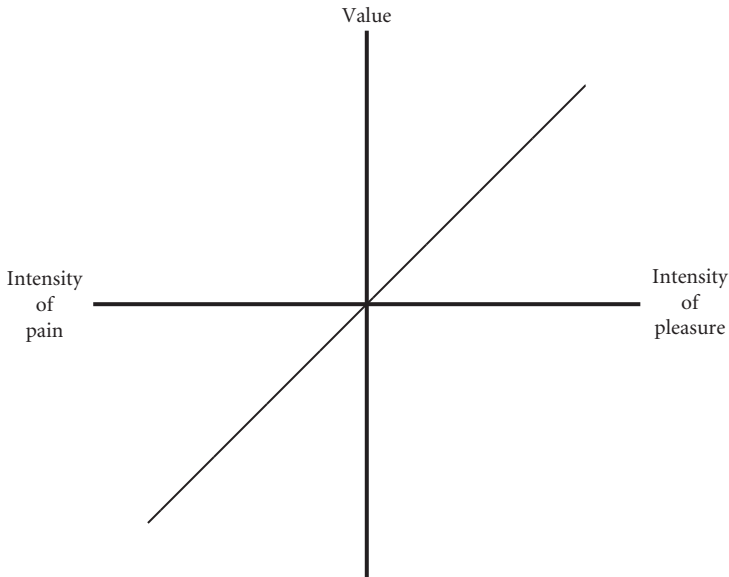


FIGURE 7.1 *Pleasure-Pain Symmetry*

states.) Now, the symmetry view will be false so long as there is one case where a pain is more evil than its corresponding pleasure is good, but I take Moore and Mayerfeld to assert a stronger *pairwise asymmetry* thesis:

For any intensity n , a pain of intensity n is more evil than a pleasure of intensity n is good.

And the simplest view that captures this thesis is represented in Figure 7.2, which still has straight lines but with different slopes above and below the horizontal axis. If the slope for pain is twice as steep as for pleasure, a pain of a given intensity is always twice as evil, considered just hedonically, as a pleasure of that intensity is good. And the basis of the pairwise asymmetry is a claim about the values of increments of pleasure and pain that I call the *marginal-value* claim:

For any intensity n , the difference in evil between n and $n + 1$ units of pain is greater than the difference in goodness between n and $n + 1$ units of pleasure.

Given a fixed initial intensity, the evil of an additional unit of pain is always greater than the goodness of an additional unit of pleasure.

There is a skeptical objection to this view. Moore and Mayerfeld assume, as Bentham and Sidgwick also do, that we can compare the intensities of pleasures and pains independently of assessing their values. But an objector may challenge this assumption, saying the claim that a pain is more intense than some pleasure merely says the pain is more evil than the pleasure is good, without pointing to some independent psychological fact that makes it so. There is no such fact, she

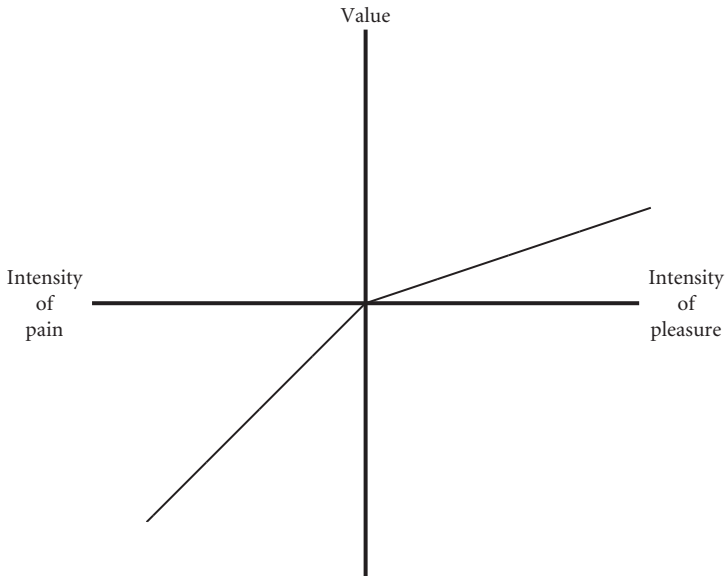


FIGURE 7.2 *Marginal-Value Asymmetry #1*

may argue, and thus no possibility of nonevaluative comparisons of pleasures and pains. If so, the difference between symmetry and asymmetry views about them disappears.⁵

This objection raises a more general issue. To formulate a pairwise asymmetry thesis about any pair of values we must be able to compare these values nonevaluatively, or identify instances of them as equal in some way that is neutral about their comparative worth. Only then can their relative values be a further issue.

For some pairs of values this nonevaluative comparison is unproblematic. To compare virtue and vice, for example, we must be able to equate a compassionate desire of intensity n to relieve a given pain with a malicious desire of the same intensity n to inflict that pain, and we can do that using the familiar economists' measures of intensity of desire. For desert and undesert, we must equate a virtuous person's enjoying pleasure of value n with his suffering pain of disvalue n , and we can do that using whatever our method is for comparing the values of hedonic states. But in other cases such comparisons do not seem possible. The question whether beauty of degree n is more good than ugliness of degree n is evil seems meaningless, since there are no comparisons of beauty and ugliness independent of comparisons of their values. Likewise for equality and inequality, or, since equality does not admit of degrees, for increases in inequality at high and low

⁵ See, e.g., James Griffin, "Is Unhappiness Morally More Important than Happiness?" *Philosophical Quarterly* 29 (1979): 47–55.

levels of inequality.⁶ But what about pleasure and pain? Do they allow nonevaluative comparisons?

That it is for this pair that symmetry issues have most often been discussed suggests that many philosophers think they do allow such comparisons, and I share this view.⁷ Surely few would extend the skeptical objection to pleasures and pains taken separately. The claim that the pain of being tortured is more intense than that of being lightly pinched does not say only that the former is worse; it points to an independent psychological fact that makes it so. Likewise for the claim that one of two pleasures is more intense. It does not follow that we can make nonevaluative comparisons across the pleasure-pain divide. We can compare temperatures with temperatures and weights with weights, but it is senseless to ask whether a fire is hotter than a stone is heavy. But pleasure and pain are not like temperature and weight. They are not completely unrelated states but states of the same kind, namely hedonic states, and they are such because the painfulness of the one is the contrary of the other's pleasantness. This does not prove that the two can be nonevaluatively compared, but it does suggest that this may be possible and it seems to be something we do.

If asked whether the pain of being tortured is more intense than the pleasure of eating a jellybean, surely we can say yes and not mean just that the pain is more evil; the same is true if we are asked about the pleasure of orgasm and the pain of being pinched. And we can give evidence for these claims. Within the categories of pleasure and pain, mild feelings make only minimal demands on our attention. We can experience the pleasure of eating a jellybean or the pain of being pinched while simultaneously feeling many other sensations, engaging in other activities, and so on. As a hedonic state becomes more intense, however, it becomes more importunate, drawing more attention to itself and starting to disrupt other activities; this can happen with orgasm on the one side and torture on the other. And we can use this fact to make pleasure-pain comparisons. If the pain of being tortured forces itself more on our attention than the pleasure of eating a jellybean, for example, that is evidence that the pain is more intense. I am not suggesting that the intensity of a pleasure or pain just is the demand it makes on our attention; the latter is only an expectable effect of the former. But as an effect it can be used to compare them in a nonevaluative way.⁸

I do not claim that these points decisively answer the skeptical objection, or expect them to persuade every reader. But for the larger purposes of this chapter it

⁶ On the multiple role of evaluations in measures of inequality, see Larry S. Temkin, *Inequality* (New York: Oxford University Press, 1993).

⁷ Mayerfeld defends the view in *Suffering and Moral Responsibility*, pp. 74–83.

⁸ It is no part of this view that the psychological facts about pleasure and pain mirror each other exactly. In particular, it is possible and even plausible that the most intense pain humans can experience is more intense than their most intense pleasure. If so, there is, asymmetries aside, a stronger reason to prevent the most intense possible pain than to promote the most intense possible pleasure. What the pairwise asymmetry grounds is a stronger reason to relieve a pain of a given intensity than to promote a pleasure of equal intensity.

is not essential that they do so. The pleasure-pain pair is just one of several for which value asymmetries are possible, and if I discuss it first it is mainly because it is the simplest, making the relevant types of asymmetry easiest to see. Those still moved by the objection should therefore suspend it temporarily, and let the pleasure-pain case illustrate, if only hypothetically, possibilities that arise in more complex ways for other, more readily comparable values.

Though it affirms a pairwise asymmetry, the view in Figure 7.2 does not capture everything either Moore or Mayerfeld says. To begin with Mayerfeld, he supplements his first claim about happiness and suffering with the further claim that it is disproportionately more important to relieve more intense pains, because they are disproportionately more evil. If we can reduce one person's suffering from 10 units to 9 or another's from 3 units to 1, he argues, we should do the former; though the resulting reduction in pain will be smaller, the reduction in hedonic evil will be greater.⁹ This second claim of Mayerfeld's is also intuitively appealing, but it requires that the evil of increments of pain not be constant, as in Figure 7.2, but increase. And this in turn requires the straight line below the horizontal axis to be replaced by a curve whose slope gets steeper as one moves to the left, as in Figure 7.3. There is no hint of this second claim in Moore's discussion, but he implies a complementary claim about pleasure that Mayerfeld never mentions. Moore's view that pleasure "has at most some slight intrinsic value" could not be true if increments

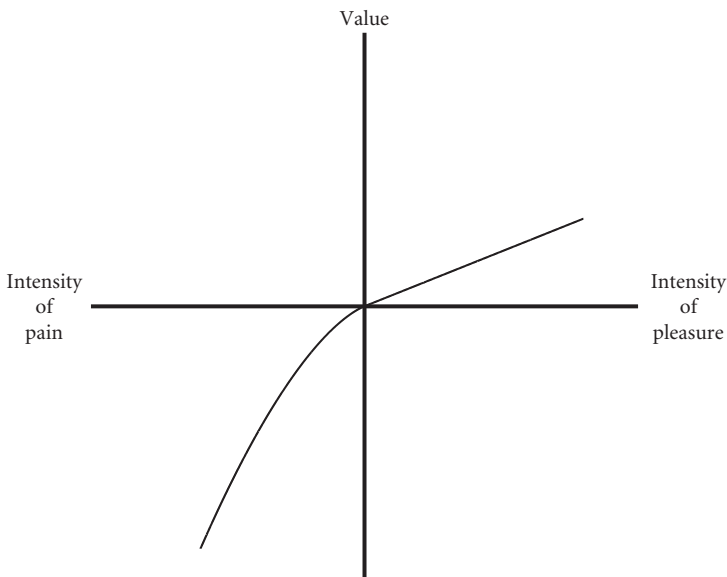


FIGURE 7.3 *Marginal-Value Asymmetry #2*

⁹ Mayerfeld, *Suffering and Moral Responsibility*, pp. 134–36.

of pleasure had constant value, as in Figure 7.2, since then a sufficiently intense pleasure could have as much value as one likes. His view requires there to be an upper bound on the value of pleasure, which in turn requires the goodness of increments of pleasure to diminish, as on Mayerfeld's view the evil of increments of pain increases. And in a pre-*Principia* article Moore explicitly mentions the possibility of diminishing marginal value for pleasure.¹⁰ If it is added to Mayerfeld's second claim, the result is the view represented in Figure 7.4, which has a smooth curve running from the bottom left quadrant to the top right, with the curve's slope getting progressively shallower as it rises to the right. This view still captures the pairwise asymmetry on a marginal value basis, and in fact makes that asymmetry stronger. But in doing so it also expresses a version of what Derek Parfit calls the "priority view," which always gives some priority to improving the condition of the worse off, interpreted here in terms of pleasure and pain. Whenever one person enjoys less pleasure or suffers more pain than another, it is better to give a fixed benefit to the first person, because the value of that benefit will be greater. The priority view is usually discussed in connection with egalitarian views about distributive justice, where it is contrasted with views that value the relation of equality as such. Here we have arrived at it by combining different attractive claims about pleasure-pain asymmetry.¹¹

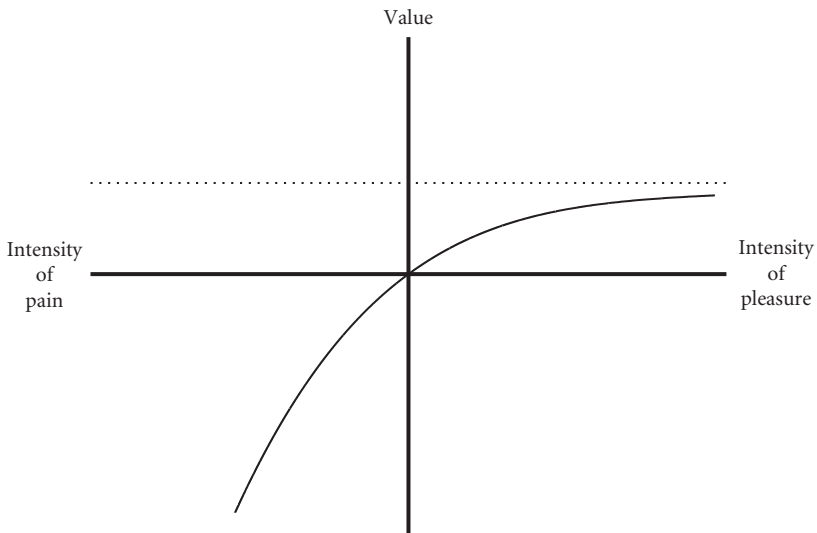


FIGURE 7.4 *Marginal-Value Asymmetry #3*

¹⁰ G. E. Moore, "Mr. McTaggart's Ethics," *International Journal of Ethics* 13 (1903): 358.

¹¹ The priority view is often formulated in whole-life terms, so it tells us to prefer benefits to those whose level of well-being has been lower through time. This can mean, contra the view in Fig. 7.4, that it is better to relieve a lesser present pain of someone who has suffered more in the past than a greater

Though still affirming a pairwise asymmetry on a marginal value basis, the priority view in Figure 7.4 implies a second asymmetry thesis, which I call the *limit asymmetry* thesis:

There is some intensity n such that a pain of intensity n is more evil than any pleasure could be good.

Because it places an upper bound on the goodness of pleasure but none on the evil of pain, the priority view allows that an intense pain can exceed in disvalue any possible value in a pleasure. This is a second way in which pain can be a greater evil than pleasure is a good: not only are its instances always more evil in equal-intensity comparisons, but they can reach heights of evil greater than any goodness possible for instances of pleasure.¹²

The priority view in Figure 7.4 supplements each of Moore's and Mayerfeld's views with a claim its author does not consider—in Moore's case about the increasing marginal evil of pain, in Mayerfeld's about the diminishing marginal goodness of pleasure—and it may give the best possible grounding of the pleasure-pain asymmetry both embrace. But the grounding of this asymmetry remains the marginal value claim, and there is an alternative possibility: one can also generate a pairwise asymmetry while rejecting the marginal value claim and holding that the values of increments of pleasure and pain are always the same. I will now explore this second strategy; though not plausible for pleasure and pain, it will prove attractive for other values.

Return to the symmetry view in Figure 7.1. The second way to ground a pairwise (though not a limit) asymmetry is to shift the single line down the graph, so it cuts the vertical axis below the origin, as in Figure 7.5. The resulting view treats increments of pleasure and pain as equal in value, thereby rejecting the marginal value claim and treating the duties to relieve pain and promote pleasure as equal in strength. But it still generates a pairwise asymmetry. If the line is shifted down, say, 2 units, then a pain of 4 units of intensity has value -6, while a pleasure of 4 units of intensity has value +2. And the asymmetry's basis is now a different, *downshift* claim:

For any intensity n , the goodness of a pleasure or evil of a pain of intensity n is $n - a$, where a is some positive number.

present pain of someone who has been happier. But the priority view can also be applied at single times. Then it says that in so far as our acts will have effects at time t , our priority should be improving the condition at t of those whose condition at t would otherwise be worst. And this single-time priority view comes close to a Fig. 7.4 asymmetry view that considers not discrete pleasures and pains, of which a person may have several simultaneously, but his overall hedonic condition at a time, which is in effect what Mayerfeld's conception of happiness and suffering does (*Suffering and Moral Responsibility*, ch. 2). So while not close to every priority view, the Fig. 7.4 asymmetry view is close to one.

¹² Note that, unlike the pairwise asymmetry discussed above, this limit asymmetry does not require non-evaluative comparisons between pleasures and pains. Even if such comparisons are impossible, some pains can be more evil than any pleasure is good.

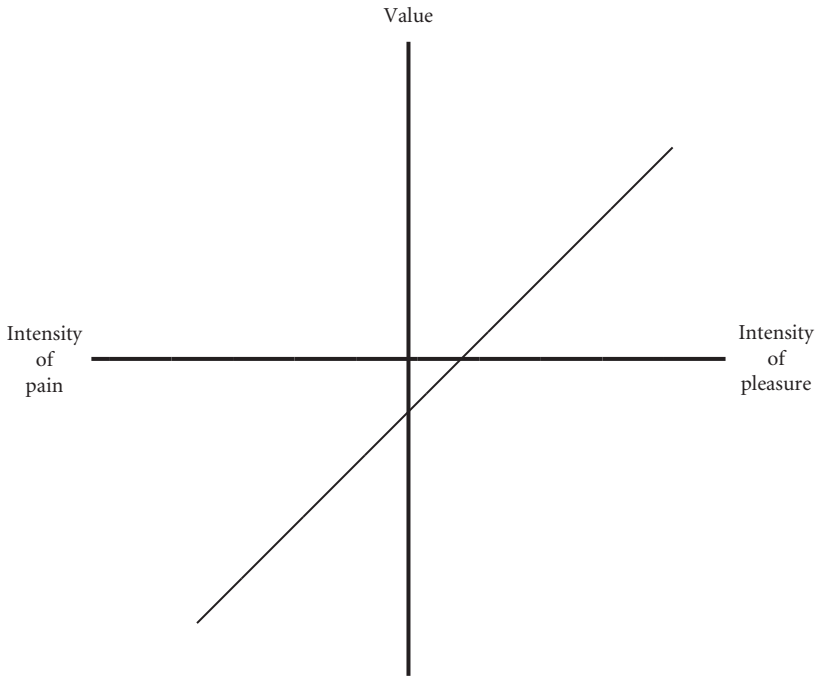


FIGURE 7.5 *Downshift Asymmetry*

An especially salient implication of this claim is that the point of hedonic neutrality, which had the value 0 in Figure 7.1, now has the value $-a$, because the line cuts the vertical axis at $-a$. But the downshift makes a similar adjustment to the value of every hedonic state, with the result that a pain of a given intensity is always $2a$ units more evil than an equally intense pleasure is good (at least when the pleasure is sufficiently intense to be good).¹³

This view, originally proposed by Gregory Kavka and generating what is now called “critical-level” utilitarianism,¹⁴ is sometimes said to avoid Parfit’s “repugnant conclusion” objection to total utilitarianism. If total utilitarianism is correct, then for any world in which billions of people enjoy ecstatic happiness, there is another world that would be better even though in it people’s lives are barely above hedonic neutrality; if there are enough such lives, the sum of goodness they contain will be greater than in the first world.¹⁵ The critical-level view avoids this implication by

¹³ This downshift too is possible without non-evaluative comparisons. So long as there is a non-arbitrary point of hedonic neutrality, as few who question such comparisons deny, both that point and pleasures only slightly above it can have negative value.

¹⁴ Gregory S. Kavka, “The Paradox of Future Individuals,” *Philosophy and Public Affairs* 11 (1982): 93–112. Critical-level utilitarianism has been defended more recently in a series of articles by Charles Blackorby, Walter Bossert, and David Donaldson; see, e.g., “Critical-Level Utilitarianism and the Population-Ethics Dilemma,” *Economics and Philosophy* 13 (1997): 197–230.

¹⁵ Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 387–90.

giving lives above neutrality but below the critical level a negative value, but its success here is surely limited. Unless it sets the critical level implausibly high, it implies, if not quite the original repugnant conclusion, then one very similar to it, in which lives in the second world are only slightly above neutrality. And it faces an even more decisive objection. Imagine a hedonically horrible world, in which billions of people suffer excruciating pain. If the critical-level view is correct, there is another world that would be worse even though everyone in it enjoys positive happiness. If their lives are below the critical-level a , their lives have negative value; and given enough such lives, the sum of evil in them is greater than in the horrible world. But surely it is absurd to say that a world containing only happiness could be worse than one in which everyone suffers excruciating pain.¹⁶ And in fact these difficulties only reflect the deeper fact that the critical-level view has no philosophically credible rationale: specific objections aside, what reason is there to believe that positive happiness is other than positively good?

I conclude that the downshift is not plausible for the hedonic values of pleasure and pain, but it is so for other intrinsic values such as virtue and vice. My understanding of virtue differs from that found in ancient philosophy or contemporary virtue ethics, though it was shared by philosophers around the turn of the twentieth century such as Hastings Rashdall, Franz Brentano, Moore, and W. D. Ross. It holds that virtue consists largely in morally appropriate attitudes to independently given goods and evils, and vice in inappropriate attitudes to them. If another person's pleasure is good, then the positive attitude of loving, or benevolently desiring, pursuing, and taking pleasure in, her pleasure for itself is virtuous and on that basis intrinsically good, while the negative attitude of enviously hating and wanting to destroy it is vicious and evil. Conversely, if another's pain is evil, the negative attitude of being compassionately pained by and wanting to relieve it is virtuous, while maliciously desiring or taking pleasure in it is evil. To discuss issues about value symmetry we need to know which features of an attitude determine its degree of virtue or vice, and here I am guided by two ideas. One is that there is an upper bound on the value of any virtuous or vicious attitude, so the attitude is always less good or evil than its object. Thus, my compassion for your pain is good, but not as good as your pain is evil. The second idea is that the best division of virtuous concern between two or more objects is proportioned to their degrees of value, so if x is twice as good as y , it is most virtuous to be twice as pleased by x as by y . A view that captures these ideas is represented in Figure 7.6, where the horizontal axis measures the intensity of love or hate for an object and each curve shows how, given a fixed value in its object, the value of an attitude as virtuous or vicious varies with its intensity. (Some virtuous or vicious attitudes may also have other values. For example, compassionate pain at another's pain may be good as compassionate but evil as pain, with its value on balance depending on how the two weigh against

¹⁶ See Gustaf Arrhenius, "An Impossibility Theorem for Welfarist Axiologies," *Economics and Philosophy* 16 (2000): 255–56.

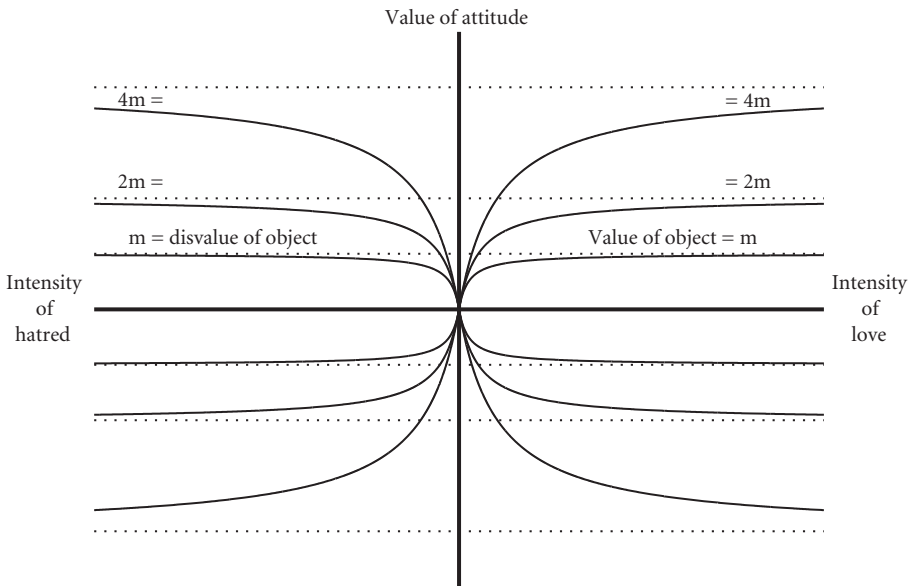


FIGURE 7.6 *Virtue-Vice Symmetry*

each other. Figure 7.6 abstracts from these other values and considers the values of attitudes only as instances of virtue or vice.) The curves for attitudes to good objects run from the bottom left to the top right, since hatred of these objects is evil and love of them good, while those for evil objects run from the top left to bottom right. And the shapes of the curves satisfy the two demands of boundedness and proportionality.¹⁷

The view in Figure 7.6 treats virtue and vice symmetrically, so that for any object a virtuous attitude to it of intensity n is exactly as good as a vicious attitude of intensity n to it is evil. It does so in part because its curves pass through the origin, so the neutral attitude of indifference to a good or evil always has zero value. But this is morally questionable, and in a way that makes a downshift positively attractive. Intuitively, indifference to another's pain is not just not good but evil; it is callousness, and callousness is a vice rather than just the absence of a virtue. Similarly, having no desire for achievable goods is sloth or apathy, which is likewise a vice rather than merely not a virtue. So here there is a positive reason to shift the curves down so they cut the vertical axis below the origin, as in Figure 7.7. More specifically, there is a reason to shift the curves for attitudes to greater goods and evils further down than those for attitudes to lesser goods and evils, so indifference to the former is a greater evil and the minimal intensities of concern for them needed for positive value are likewise greater. And making this downshift results in

¹⁷ I give a fuller account of these issues in *Virtue, Vice, and Value* (New York: Oxford University Press, 2001), ch. 3.

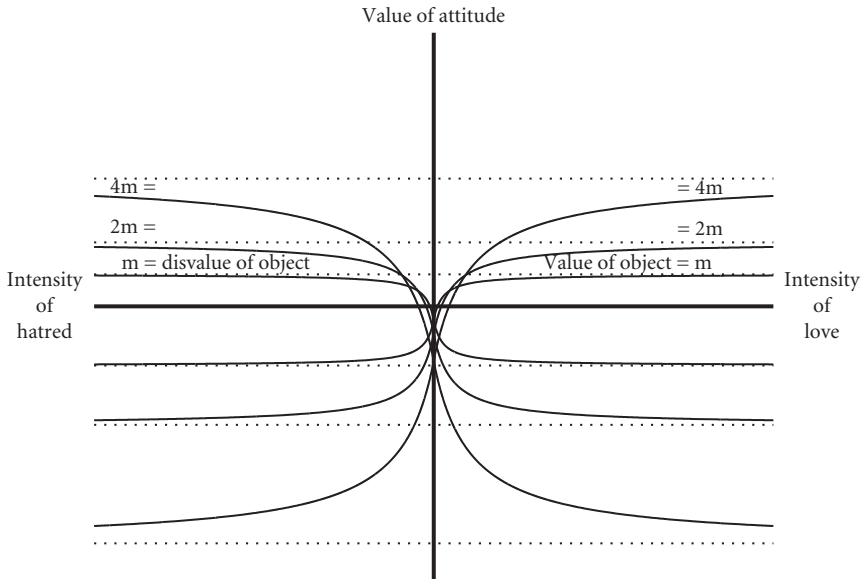


FIGURE 7.7 *Virtue-Vice Downshift Asymmetry*

an asymmetry whereby vice is a greater evil than virtue is a good in both pairwise and limit ways. Go equal distances to the left and right of the origin, and the distance down to the vice portion of a given curve is always greater than the distance up to its virtue portion. This makes a malicious desire of intensity n to inflict a given pain more evil than a compassionate desire of intensity n to relieve the pain is good. Given paired instances of malice and compassion, the former is more vicious than the latter is virtuous, or more evil than the latter is good. In addition, the lower bound on the value of a vicious attitude to an object is always further below the horizontal axis than the upper bound on the value of a virtuous attitude to it is above the axis, so a malicious desire to inflict a given pain can be more evil than any compassionate desire to relieve it is good. Supplementing the pairwise asymmetry, then, is a further limit asymmetry.

If the downshift in Figure 7.7 is uniform along each curve (as it must be to satisfy the proportionality condition), it makes not only indifference but also very weak appropriate attitudes, such as very mild compassion for great pain, intrinsically evil. This is also intuitively appealing; it seems right that feeling only mild distress at, say, the Holocaust is not just not good but evil. That too is a form, though a lesser one, of callousness. And this implication can be given a positive rationale: if what is evil is morally inappropriate attitudes, then one way an attitude can be evil is by being, while properly oriented, inappropriately weak for its object. For both these reasons, the downshift for virtue and vice does not invite a decisive objection like the one against critical-level utilitarianism. A world in which a huge number of people feel only mild distress at the Holocaust can indeed be worse than one in which a much smaller number take positive pleasure in it;

the more extensive callousness in the first world can be more evil than the less extensive malice in the second.

In Figure 7.7 the asymmetry depends entirely on the downshift and not at all on a marginal-value claim; because the curves retain their symmetrical shapes from Figure 7.6, the value of equivalent increments of virtue and vice remains the same. For this reason, the view in Figure 7.7 does not support an analogue of Mayerfeld's claim that the duty to relieve suffering is stronger than the duty to promote happiness. On the contrary, though it makes a vicious attitude to a given object more evil than the corresponding virtuous attitude is good, it holds that the duties to reduce vice and promote virtue are equally strong. The strengths of these duties depend only on the slopes of the curves above and below the point representing indifference; if those slopes remain unchanged, as they do in Figure 7.7, so do the duties' weights. Now, one could supplement the view in Figure 7.7 with a marginal value claim, giving the curves a steeper slope below the point of indifference and pushing the lower bounds even further down. This would strengthen both the pairwise and limit asymmetries, and make the duty to reduce vice stronger than the duty to promote virtue. But though I do not have a firm view about this possibility, I do not find it that attractive. I see no reason why intensifying virtuous love for an object should be less good than intensifying vicious hatred of it is evil, and I will therefore retain the simpler view in Figure 7.7. Doing so makes for an interesting contrast with the case of pleasure and pain. For those hedonic values the most attractive asymmetry rests only on the marginal value claim and the alternative downshift basis is not plausible, while for virtue and vice the best asymmetry depends only on the downshift and not on marginal values. Of the two possible bases of asymmetry in value, only one suits the one pair and only the other suits the other. And this contrast can be extended, since there is a further pair of values for which both grounds of asymmetry are plausible.

This pair contains the good of desert, which I will understand more specifically as moral desert, or getting what one deserves on the basis of one's moral qualities, and its contrary undesert, or getting the opposite of what one deserves. Given these values, it is good if virtuous people enjoy pleasure or vicious people suffer pain, and bad if the virtuous suffer or the vicious are happy. These desert values parallel virtue and vice in many ways—for example, by making a positive response to a positive value, this time the reward of happiness for the good of virtue, intrinsically good, and a negative response to that same value evil. They are also governed by similar ideas of boundedness and proportionality. The goodness of rewarding virtue or punishing vice is always less than the goodness of the virtue or evil of the vice; it is not better to have vice and its deserved punishment than to have no vice at all. And the best division of rewards or punishments among people is proportioned to their degrees of virtue or vice, so that, for example, those who are twice as virtuous enjoy twice as much pleasure. But there is an important difference between the two pairs. In Figures 7.6 and 7.7 the slopes of the virtue curves, while diminishing, always remain positive, so a more intense love of a good is always intrinsically better. A more intense love may be instrumentally worse if it prevents one from having other,

more valuable attitudes, but in itself it is always preferable. But it would not be plausible to take a similar view about desert. In particular, it would not be plausible to say that if a person is vicious, it is always better from the point of view of desert if he suffers more pain. Desert values demand a different structure, whereby for any degree of virtue or vice there is a specific amount of pleasure or pain that it ideally deserves and that has most value, and where amounts above or below that ideal are less good and can even be evil. A desert view with this “peak” structure is represented in Figure 7.8. Here the horizontal axis measures the amount of pleasure or pain a person experiences while the vertical axis measures how good or evil this is just as a matter of desert. (The resulting desert values must again be weighed against other values, such as hedonic ones, to determine the value on balance of his enjoying the pleasure or suffering the pain. Thus, the pain of a vicious person, while good as deserved, is bad as pain.) The topmost points on the different curves represent the ideal rewards and punishments for different degrees of virtue and vice, and the curves’ shapes again incorporate the demands of proportionality.¹⁸

Figure 7.8 already contains the first or pairwise asymmetry, on a marginal value basis. It follows from the proportionality condition that the downward slopes on the inside of a peak, the side closer to the vertical axis, get steeper until they cut the vertical axis. Below that point they have to get shallower, but in Figure 7.8 the rate of change in slope below the cut is much slower than the rate of change above it, so they stay close to the vertical longer. This implies that if we go equal distances to the left and right of the origin, the distance up to the desert portion of a given curve is always less than the distance down to its undesert portion. If a virtuous person ideally deserves, say, 4 units of happiness, his enjoying that happiness is less good from the point of view of desert than his suffering 4 units of pain is evil. And the view in Figure 7.8 can be supplemented by a downshift, if we shift the curves so they cut the vertical axis below the origin. In the case of virtue, the neutral response of being indifferent to goods or evils was plausibly not just not good but evil. Here it is likewise plausible that a virtuous person’s getting no reward but experiencing neither pleasure nor pain is positively evil in desert terms, as is a vicious person’s suffering no punishment. Doing nothing to meet the demands of justice, or giving no response to merit or demerit, is a positive injustice. A graph that incorporates this downshift is given in Figure 7.9; with two bases for the pairwise asymmetry, it makes that asymmetry even stronger. In a book on punishment A. C. Ewing writes,

¹⁸ For a similar peak structure (though without proportionality) see Shelly Kagan, “Equality and Desert,” in Louis P. Pojman and Owen McLeod, eds., *What Do We Deserve? A Reader on Justice and Desert* (New York: Oxford University Press, 1999), pp. 298–314. In Fig. 7.8 the “degree of virtue/vice = 0” line represents the desert-value of happiness and suffering for people who are neither virtuous nor vicious, or whose characters are at the neutral point between the two. This line has a peak at the origin, implying that what is best is these people’s experiencing neither happiness nor suffering, though their doing so has neutral value, and that their experiencing either happiness or suffering has negative value. This may seem harsh; should their experiencing happiness or suffering not also have neutral value? But it is required by the peak structure plus proportionality.

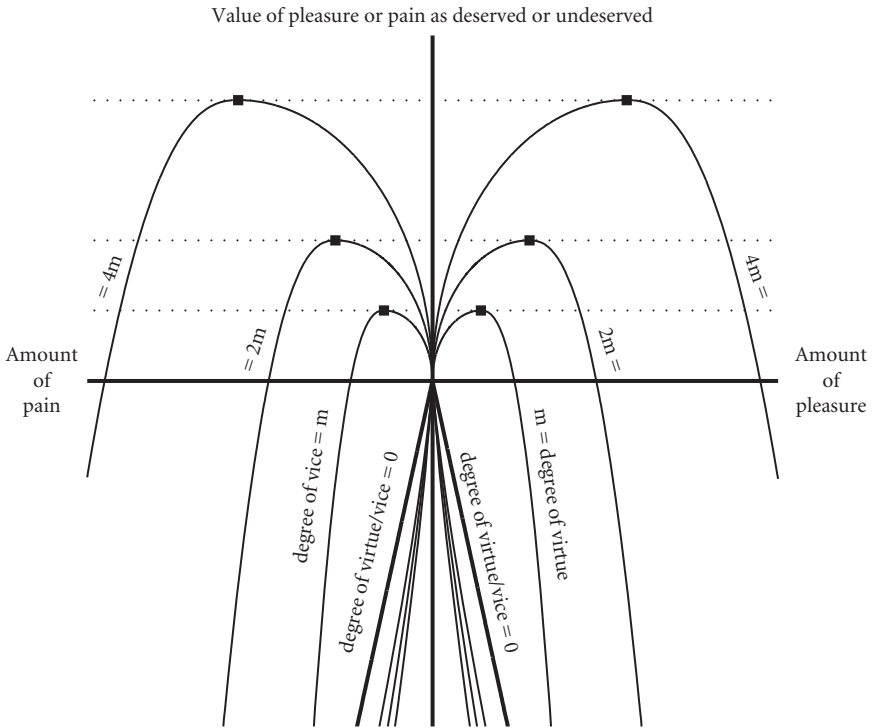


FIGURE 7.8 Desert Marginal-Value Asymmetry

“When I look at the two, injustice in punishment seems to me a *very much greater intrinsic* evil than justice is a good, especially if the injustice consists in punishing somebody for an offence of which he is not guilty or in excessive severity.”¹⁹ Figure 7.9 gives two different grounds for this intuitively appealing claim.

As in the case of virtue, the downshift-based asymmetry does not make the duty to prevent undesert stronger than the duty to promote desert, because it does not on its own change the curves’ slopes. But the marginal value-based asymmetry does have this effect, and in so doing connects with a familiar thesis about criminal punishment—namely that the state should be more concerned about not punishing the innocent than about punishing every one of the guilty. As Sir William Blackstone put it, “Better that ten guilty persons escape than that one innocent suffer.”²⁰ This thesis can be grounded in deontological principles, if the state has a stronger duty not to *do* what punishes the innocent than not to *allow* the guilty to go free. But it can also be supported by claims about value if the incremental evil

¹⁹ A. C. Ewing, *The Morality of Punishment* (London: Kegan Paul, Trench, Tubner & Co.: 1929), p. 109.

²⁰ Sir William Blackstone, *Commentaries on the Laws of England* (Chicago: University of Chicago Press, 2002), Bk. 4, ch. 27, p. 352; for a survey of variants on Blackstone’s formula, see Alexander Volokh, “*n* Guilty Men,” *University of Pennsylvania Law Review* 146 (1997): 173–216.

of punishing the innocent is greater than the incremental good of punishing the guilty. Then, even apart from deontological considerations, punishing the innocent does more to violate the demands of justice than punishing the guilty does to satisfy them. The criminal desert relevant to Blackstone's thesis differs in several respects from moral desert, for example, by concerning specific criminal acts rather than overall moral virtue. But it is still governed by a proportionality condition, so the best division of punishments among crimes is proportioned to their degrees of seriousness, and it still supports a downshift, since it remains plausible that a criminal's escaping punishment is not just not good but evil. An attractive axiological view of criminal desert therefore supports Blackstone's thesis and the legal procedures it justifies, such as the presumption of innocence in criminal trials. It also supports a subtle consequentialist account of the morality of punishment proposed by Ewing. In this account considerations of desert play only a modest positive role. Though punishing wrongdoers is an intrinsic good, it is a comparatively minor one and does less to justify legal punishment or to fix its optimal severity than do considerations of deterrence and moral reform.²¹ But

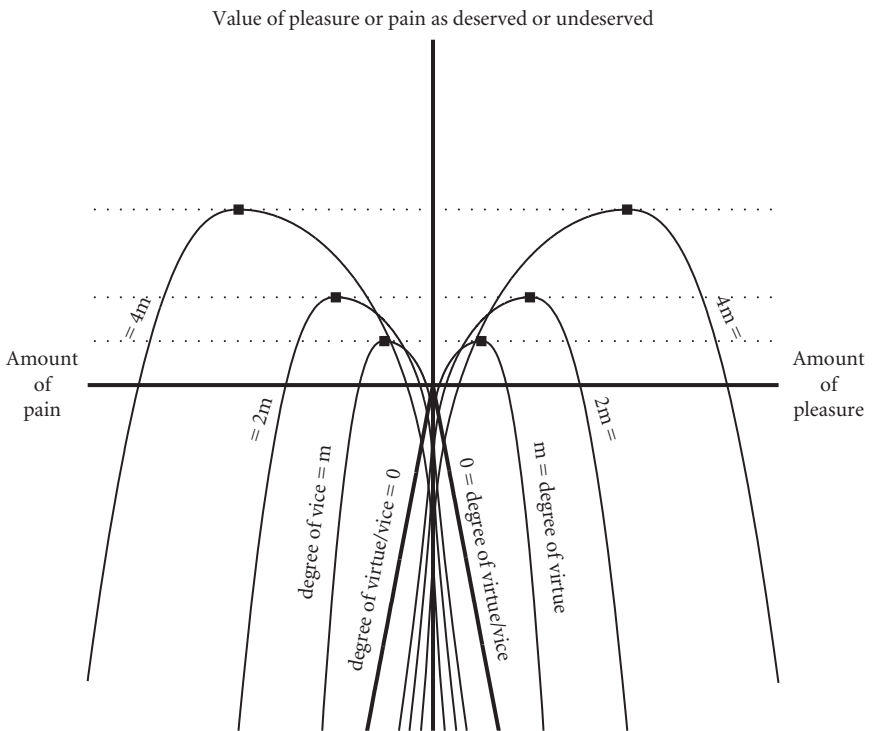


FIGURE 7.9 *Desert Double Asymmetry*

²¹ The last claim is not true for excessive punishment, which on Ewing's view can be a great evil. It is when deterrence and reform favor milder punishments than retribution that they outweigh it; when they favor harsher ones, they do not.

desert is much more important on the account's negative side, which says punishment may normally be inflicted only on wrongdoers and not on the innocent. Because punishing the innocent is a great intrinsic evil, retribution counts strongly against such punishment and will often forbid it when deterrence and reform do not. So the asymmetry between desert and undesert leads to an asymmetry of justifying roles: retribution counts only modestly in favor of punishing the guilty, which is primarily justified on other grounds, but does much more to forbid punishing the innocent.

Figures 7.8 and 7.9 also contain the second or limit asymmetry. If there is an upper bound on the goodness of desert but no lower bound on the evil of undesert (as there cannot be given proportionality), then whatever a person's degree of virtue or vice, his getting the opposite of what he deserves can be more evil than his getting what he deserves can be good. But in this case the limit asymmetry has a more specific basis. The value of desert is not just bounded but can be fully achieved, if a person gets exactly the happiness or suffering he merits. The relevant desert-goodness is then complete, in the sense that it cannot be improved. This was not the case with pleasure or virtue; since their curves never reached a peak, their instances could always in principle be made better. While those values are not fully achievable, desert is. And this provides a distinctive basis for asymmetry that is also found in a good like equality, understood as a relation between people's levels of, say, happiness that is valued as a relation and is opposed to an evil of inequality. Like desert, equality can be fully achieved, if people's levels of happiness are exactly the same. But its contrary, inequality, can in principle increase without limit. Assuming a finite value for equality, therefore, there can be unequal distributions among people that are more evil than any equal distribution among them can be good, so, as in the case of desert, injustice is in the limit sense a greater evil than justice is a good.²² These asymmetries may echo an ancient idea due to the Pythagoreans and discussed sympathetically by Plato and Aristotle, an idea that associates good with "limit" and evil with the "unlimited." As Aristotle says, "evil belongs to the class of the unlimited, . . . and good to that of the limited."²³ The thought here is not just that desert and equality are subject to limits, in that their value has an upper bound. It is that they themselves involve a limit, or a mathematical relation that can be completely achieved. On the Pythagorean view this makes desert and equality good, in contrast to opposites that can increase without limit. But in an implication the Pythagoreans may have found less welcome, it also makes their value less, in the limit sense, than their opposing intrinsic evils.

²² Note that, as before, a limit asymmetry does not require nonevaluative comparisons, which I have said are not possible for equality and inequality.

²³ Aristotle, *Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984), 1106b28–30. The remark comes when Aristotle is introducing his doctrine of the mean.

Let me summarize. I have examined three main pairs of values and in each found either or both of a pairwise and a limit asymmetry. The bases of the asymmetries have been different: for pleasure and pain just the marginal value claim, for virtue and vice just the downshift, and for desert and undesert both. But there is an obvious pattern to the asymmetries. In each pair it is the evil that is greater than the good, or the evil that is morally more potent. Contrary good-favoring asymmetries can certainly be formulated, but they do not have much intuitive appeal. Surely no one would say it is more important to increase the happiness of the very happy than to decrease the suffering of the miserable, or that the neutral attitude of indifference to another's pain is positively good. In all these pairs the intuitive pressure is to accentuate the negative. Nor should this be surprising, since there are other parts of morality where negative considerations are more potent.

Consider a deontological morality that sometimes forbids acts that have the best outcome, and imagine that, like consequentialism, it formulates its principles by reference to good and evil states of affairs rather than to some alleged Kantian value in persons. This morality may use either or both of the distinctions between doing and allowing and between intending and merely foreseeing, but these distinctions are engaged only by the production of evils and not by the production of goods. If I can relieve five people's pain by directly causing another person pain, a deontological morality may say it is wrong for me to do so. But if I can either directly cause one person's pleasure or allow someone else to cause pleasure for five, the same morality will say I should prefer the pleasure of the five. While it is more objectionable to actively cause evil, it is not usually more creditable to actively cause good. A similar point applies to intention and foresight. It may by deontological lights be wrong to intend one person's pain as a means to relieving five other people's, but right to prefer an act that merely foresees pleasure for five to an act that intends it for one.²⁴ So in a deontological context evil is morally more potent than good because it engages the central deontological distinctions where good does not. The asymmetry here is not exactly analogous to the ones we have discussed, because it is not between evil and good as such; it is between changes that make a situation worse and changes that make it better. If to prevent five people from having their happiness reduced I must directly reduce the happiness of one, a deontological morality may forbid this act even though it involves only the good of happiness, and likewise if I intend the reduction. So what differentially engages the deontological distinctions is not good and evil as such, but changes in the values of states of affairs for the worse and for the better. Nonetheless, there is a fundamental asymmetry in these moralities that makes a kind of negative effect more potent than its corresponding good one.

²⁴ That the intention-foresight distinction is engaged only by the production of evils and not by that of goods is noted in Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), p. 181.

The pattern whereby in three pairs of values the evil is greater than the good has implications for several traditional philosophical issues. Consider the nineteenth-century debate between optimists and pessimists about whether the world is on balance good or on balance evil. The asymmetries we have identified strengthen the case for pessimism; since it now takes more pleasure or virtue to outweigh a given quantity of pain or vice, the result of the weighing, given a fixed set of facts, is more likely to be negative. The asymmetries also affect the theological problem of how an all-powerful, perfectly good God could create the evil we see around us. It again takes more pleasure to outweigh a given quantity of pain in God's creation, and if God gave humans the possibility of either benevolence or malice, then, assuming an initially equal probability of each, he did something whose expected moral value was negative. The obvious question, however, is whether this pattern of evil-favoring asymmetries is universal or whether in some pairs a good is greater than its corresponding evil. I will discuss this question in the last part of this chapter, after first considering some different asymmetries between positive and negative forms of the same good or evil.

Return to Figures 7.6 and 7.7. Though they differ in how they relate virtue and vice, they are both symmetrical around the vertical axis, and the views they express therefore treat positive and negative forms of virtue and vice as equal in value. Imagine that a given pleasure is exactly as good as a given pain is evil. (For Bentham and Sidgwick this will mean the two are equally intense, for Moore and Mayerfeld that the pleasure is to a specified degree more intense.) Figure 7.7 makes a benevolent pleasure of intensity n in the pleasure exactly as good as a compassionate pain of intensity n at the pain: the n -point on the curve in the top right quadrant is exactly as high as the n -point on the comparable curve in the top left. It likewise makes a malicious pleasure of intensity n in the pain exactly as evil as an envious pain of intensity n at the pleasure. Should we retain this feature of the graphs or supplement our up-down asymmetry with a further, left-right asymmetry?

The answer depends on whether positive and negative forms of virtue should be proportioned to each other, and in my view they should. If it is disproportionate, and more specifically selfish, to prefer a minor good for oneself to great goods for other people, surely it is equally disproportionate to prefer avoiding a minor evil for oneself to securing great goods for others, and likewise disproportionate to prefer a minor good for oneself to preventing great evils for others. The demands of proportionality apply not only within the categories of virtuous love and hatred but also across them, so an ideally virtuous person divides all his concerns, both positive and negative, in proportion to their objects' values. And this requires symmetry around the vertical axis: if virtuous love and hate are to be balanced proportionally, the functions determining their values must mirror each other, as in Figure 7.7.

A similar issue arises about desert. While Figures 7.8 and 7.9 make undesert a greater evil, they too are symmetrical around the vertical axis and so make positive and negative desert equally good. If one person's virtue is exactly as good as

another's vice is evil, then the first's getting n units of happiness is exactly as good as the second's suffering n units of pain. Should we retain this symmetry, as we retained one for virtue?

The answer again depends on whether there are proportionality demands between positive and negative forms of goodness, and here I am less persuaded. Do we think society should balance its rewards and punishments, so that, say, a Nobel Peace Prize makes a person exactly as happy as life imprisonment makes one suffer? I do not think we have any such concern. We care that rewards be proportioned to rewards and punishments to punishments, but seem not to care about proportionality between the two categories. This leaves room for positive-negative asymmetries about desert, and here two seem attractive. I think many who value desert will say there is a stronger demand to punish the vicious than to reward the virtuous, so failing to do the former is a greater failing in justice, or involves a greater loss of value.²⁵ To reflect this view, we can make the peaks on the curves to the left of the vertical axis higher, so they represent greater positive value, and also make those curves cut the vertical axis further below the origin, so ignoring that value is worse. (This will be another case where a broadly negative value is more potent: returning evil for evil will be more important than matching good with good.) Second, compare the slopes of the curves on either side of a peak. In Figures 7.8 and 7.9 the slopes are always steeper outside the peak, so getting more happiness or suffering than one deserves involves a greater loss of value than getting less. This is certainly attractive for negative desert. Just as the state should be more concerned not to punish the innocent than to punish all the guilty, so it should be more concerned not to punish the guilty too severely than not to punish them enough. But the parallel view is much less plausible for positive desert, where giving an excessive reward seems less bad than giving an insufficient one.²⁶ If so, the shapes of the curves should differ on the two sides of the vertical axis, being steeper on the outside on the left, for negative desert, and on the inside on the right.²⁷

²⁵ Ewing denies this, suggesting that rewarding the virtuous is just as good as punishing the vicious; see *The Morality of Punishment*, p. 147.

²⁶ This is certainly Kagan's view, and is captured in the "bell-motion" feature of his desert graphs; see "Equality and Desert," p. 301.

²⁷ This asymmetry could in principle be taken further. Some may say that, while a peak structure is necessary for negative desert, it is never less good in desert terms if a virtuous person is happier; on the contrary, it is always better. To capture this view, the desert curves would have to have entirely different shapes on either side of the vertical axis, turning down from a peak to the left but continuing up toward a limit on the right. But this more radical asymmetry is problematic. It requires the right half of the "degree of virtue/vice = 0" line, which is below the horizontal axis in Figs. 7.8 and 7.9, to lie along that axis, implying that if a person is neither virtuous nor vicious his enjoying happiness has zero rather than negative desert-value. And while this may seem attractive in itself, it introduces a serious discontinuity into the account of desert-value. Imagine that a slightly vicious person enjoys great happiness. Given the more radically asymmetrical view, improving his virtue to the neutral point between virtue and vice makes for a very large increase in desert-value, from a possibly large negative value to zero. More specifically, it can make it better in desert terms, and perhaps even on balance, to make a small improvement in virtue if that takes one to or across the neutral point than to make larger increases

Our main question, however, is whether our initial pattern of evil-favoring asymmetries is universal, or whether there are cases where a good is greater than its corresponding evil. This is not, it will turn out, an easy question.

To pursue it, consider Robert Nozick's fantasy of an experience machine that, by electrically stimulating the brain, can give one the experience and therefore the pleasure of any activity one likes. And assume with Nozick that life on this machine would not be intrinsically best and, in particular, would not be as good as the lives we currently lead. A question that has not to my knowledge been discussed is whether life on the experience machine merely lacks goods found in ordinary life or, while containing some goods such as pleasure, also contains evils that weigh against those goods. Does machine life merely lack positive value or does it also contain some negative value? Like Nozick, I take what is problematic about life on the machine to be its disconnect from reality. People who plug in have false beliefs about the world and their place in it; they think they are, say, climbing Mt. Everest when they are not. Nor do they actually accomplish any goals; while they intend to climb Everest, they do not do so. So there are two pairs of values highlighted by Nozick's example. One is knowledge and its contrary, false belief, surrounding the neutral state of not having any opinion about a subject. The other is achievement and its contrary, failure in the pursuit of a goal, with the neutral state of not pursuing the goal at all. Do both these pairs contain an intrinsic good and an opposing intrinsic evil? If so, do the good and evil relate symmetrically or is one greater than the other?

Let me begin with knowledge, and consider first the kind that involves knowing one's relation to the external world, and in particular to one's immediate environment. The absence of this knowledge seems a large part of what is troubling about the experience machine: that people on it believe falsely that they are doing things rather than being electrically stimulated plays a large role in making their condition less good. But I think it involves more the presence of an evil than merely the absence of a good. If having true beliefs about one's current environment is good—which some may dispute given the extreme particularity of its subject matter—it is surely not a great good. If someone with a painful terminal illness believes correctly that he is lying in a hospital bed, the goodness that knowledge involves does not weigh heavily against the evil of his pain. So if the delusions about one's place in the world generated by the machine do weigh heavily against its pleasures, they must be a positive evil, and that seems intuitively right. Being systematically mistaken about where one is is not just not good but evil, and more evil than its contrary is good. It may be objected that having false beliefs about one's place in the world is not much worse than having no beliefs about it, or not being in a position

either below or above that point, for example, to improve from 1 unit of vice to neutrality than to improve from 5 to 1 units of vice or from 0 to 4 units of virtue. If this discontinuity is counterintuitive, that is a reason to resist the more radical left-right asymmetry and adopt at best the milder one that retains peaks on both sides of the vertical axis.

to have beliefs about it. But that only shows that the asymmetry may have a downshift basis, where a neutral state has negative value. So we seem here to have an evil-favoring asymmetry like the one for virtue and vice: knowing one's relation to one's immediate environment has at best modest positive value, but being mistaken about it has greater negative value, in part because not having an opinion also has negative value.

But now consider a different kind of knowledge, of purely external facts about the world such as scientific laws. And think of some scientist of the past who had mostly false beliefs on this topic, as Aristotle did about the basic laws of physics and biology. Aristotle's errors about these laws do not seem significantly evil, especially in comparison with his nonscientific contemporaries who had no beliefs at all about them. Nor does those contemporaries' lack of beliefs seem evil—it is just the absence of a good—so there is here no downshift. A case like Aristotle's is complicated because there were other intrinsic goods associated with his scientific activities. His beliefs about physics and biology were arguably justified by his evidence, and there may be positive value in having justified beliefs even when they are false. Moreover, his active pursuit of true scientific beliefs showed a love of scientific knowledge that is a form of virtue. But I think we can abstract from these goods and consider his false beliefs on their own, and when we do, I do not think we find them significantly evil. But having true beliefs or knowledge about scientific laws does seem significantly good. Most philosophers who have discussed the intrinsic value of knowledge have held that the best knowledge is of the most general and explanatory principles, including pre-eminently the knowledge of scientific laws.²⁸ So here we seem to have an opposite, good-favoring asymmetry: knowing scientific laws is significantly good, but the contrary state of being mistaken about them is not significantly evil.

Though there are other kinds of knowledge, this value has already proved complex. For one subject matter, that of one's relation to one's immediate environment, the evil of false belief seems greater than the good of knowledge, while for another, concerning scientific laws, the good of knowledge seems greater.²⁹ But there is here at least a partial break from our earlier pattern of evil-favoring asymmetries, in that sometimes knowledge is a greater good than false belief is evil. And at least one philosopher has affirmed this view. Making a contrast with hedonic values, though not distinguishing among subjects of knowledge, E. F. Carritt says, "Pain seems more obviously bad than pleasure is good, but knowledge more plausibly good than either ignorance or error is evil."³⁰

²⁸ See, e.g., F. H. Bradley, *The Principles of Logic*, 2nd ed. (Oxford: Oxford University Press, 1922), pp. 684–88; W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 139, 147–49; and Thomas Hurka, *Perfectionism* (New York: Oxford University Press, 1993), pp. 115–20.

²⁹ A third kind of knowledge is of one's inner states, not only one's experiences but also one's character traits and dispositions. I have no clear intuitions about whether either knowledge of or delusion about these states is a greater intrinsic value.

³⁰ E. F. Carritt, *Ethical and Political Thinking* (Oxford: Clarendon Press, 1947), p. 94.

When we turn to the second relational good, achievement, there is an even sharper break, since the contrary state of failure in pursuit of a goal seems not evil at all. Of romance Tennyson said, “’Tis better to have loved and lost/Than never to have loved at all.” Of practical endeavours we may say, similarly, “’Tis better to have sought and failed than never to have sought at all,” so failing in pursuit of a goal is if anything better than not pursuing it. This conclusion is overdetermined when the goal in question is independently good, since then pursuing it, even without success, manifests the virtue of loving the good, while declining to pursue it may be viciously indifferent. But the conclusion seems also to hold for failures to achieve intrinsically neutral goals, as in games or business. While successfully achieving a neutral goal such as a low golf score or large profits for one’s company can be significantly good, failing to achieve it does not seem significantly evil, and in particular does not seem worse than not pursuing it. Nor does this case involve a downshift, since not pursuing a neutral goal seems neutral in value rather than evil. So for this pair, and without the distinctions we found for knowledge, there seems to be a broad-based good-favoring asymmetry whereby achievement is a greater good than failure is an evil, if failure is evil at all. Together with the more restricted good-favoring asymmetry for knowledge, this provides a counterweight to the contrary asymmetries in discussions of optimism versus pessimism and the problem of evil. Whereas the pleasure-pain and virtue-vice asymmetries make the world more likely to be on balance evil and God’s creation of it harder to reconcile with his goodness, those for knowledge and achievement make it more likely to be good.³¹ It is not always evil that is more potent; in some cases good has greater weight.

This chapter does not have a grand conclusion. I have explored several possible asymmetries within pairs of intrinsic values, and while it would be exciting to report a single pattern in these asymmetries, I have not found one. There was an initial pattern of evil-favoring asymmetries in four pairs of values, but even there desert and undesert did, but virtue and vice did not, support a further left-right asymmetry between positive and negative forms of good or evil. And my later discussion, though more tentative, seemed to find some contrary good-favoring asymmetries in at least two other pairs of values. But even without a single pattern I hope the exploration has been illuminating. Our first impulse may be to relate the elements of a pair of values symmetrically, and this is certainly the simplest view. But there is no necessity for value symmetry, and often an asymmetrical view is more attractive. Moreover, there are several different possible asymmetries and several different bases they can have, with different implications for claims about the right. As Moore said, ethics would be far simpler if values always related to each other symmetrically, but the truth seems rather more complicated.

³¹ Perhaps there are other values with the same effect, such as beauty and creativity. They too may be more good than their contraries are evil, if those contraries are evil at all.

This page intentionally left blank

PART III

Individual Goods

This page intentionally left blank

Why Value Autonomy?

I. Autonomy and Goodness

Many of us think that autonomy is intrinsically good. When we imagine an ideal human life we think that its leading features must be chosen by the agent herself, and chosen from many options all fully understood. We want a person to direct her life, and to do so meaningfully. To be autonomous, on at least one understanding, is to direct oneself where different directions are possible. For many of us this autonomy is a good.

In so valuing autonomy we embrace a mild perfectionism, a mild version of the view that states and activities can be good objectively, apart from people's attitudes to them. This mild perfectionism, I would argue, is present in the most widely accepted common moralities. And it has been prominent in our political tradition. The classical liberals, Humboldt and Mill, condemn state interference in private activities in part because of its effects on citizens' autonomy. Humboldt's ideal is that each person "develop himself from his own inmost nature, and for his own sake," and he thinks this requires liberty, and what is "intimately connected" with liberty: "a variety of situations."¹ Mill says that a person "whose desires and impulses are his own," and which have been developed "by his own culture," has "a character," while one whose impulses are not his own "has no character, no more than a steam-engine has a character." Contrasting autonomy with a slavish obedience to custom, Mill says that custom

does not educate or develop in [a person] any of the qualities which are the distinctive endowment of a human being. The human faculties of perception, judgement, discriminative feeling, mental activity, and even moral

¹ Wilhelm von Humboldt, *The Limits of State Action*, trans. J. W. Burrow (Cambridge: Cambridge University Press, 1969), pp. 16, 19.

preference, are exercised only in making a choice. He who does anything because it is the custom makes no choice.

What follows for Mill are political recommendations very like Humboldt's: no interference in the private realm, and "experiments of living" to make different life options vivid for all.²

This perfectionist valuing of autonomy is not the only possible ground for liberalism, but it is intuitively appealing and has been important in our tradition. On closer inspection, however, it can seem problematic. Autonomy, I assume, involves choice from a wide range of options.³ And it can seem puzzling that this should have *intrinsic* value. Imagine that there are ten possible actions, which an agent ranks in order of preferability from one to ten. (This is the agent's ranking. We could also use our ranking, or an objective ranking, without fundamentally altering the example.) If all ten are available to her she can choose autonomously among them, or choose more autonomously than if she only had one. But what if that one were the highest-ranked action, the one she prefers to all others? Why would it be worse to lack autonomy if she still had her most favored option? Things are different, of course, if we compare having ten options with having only the lowest-ranked among them. Then having more choice is clearly better. But this need have nothing to do with autonomy. We may value the extra options just because they are better, and offer the prospect of better choice. We may, in other words, value the expansion of choice just *instrumentally*, as a means to better results. The important case for the *intrinsic* value of autonomy is the first case, where the extra options are all worse. And this case can seem puzzling. Why should it be better to choose autonomously among ten options than to have only the best among them? Why value extra possibilities when they cannot rationally be chosen?⁴

This, then, is a challenge to the value of autonomy: if free choice is intrinsically good, it should be better to have one good option and nine bad ones than to have just the good option. And why should this be so? In this chapter I will try to meet this challenge, and to vindicate what I take to be our initial intuitions about the value of autonomy. But first I must describe two views that reject the challenge, by denying that its questions need answers.

² John Stuart Mill, *On Liberty*, in *Utilitarianism, Liberty, and Representative Government* (London: J. M. Dent, 1915), pp. 115, 116, 118.

³ Let me emphasize that this is not the only conception of autonomy. On another perfectly acceptable conception, a person with limited options could count as autonomous if she viewed her options in the right way, after the right kind of reflection. This conception does not have the political implications of the one I am assuming, but it is an acceptable conception of a far from univocal concept.

⁴ Versions of these questions occur in Gerald Dworkin, "Is More Choice Better than Less?" in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds., *Midwest Studies in Philosophy* 7 (1982): 47–61, and Joel Feinberg, "The Interest in Liberty on the Scales," in his *Rights, Justice and the Bounds of Liberty* (Princeton, N.J.: Princeton University Press, 1980), pp. 30–44, though both writers are primarily interested in the instrumental value of liberty. Dworkin argues briefly, and unconvincingly, against intrinsic value on pp. 59–60.

The first view concedes that the value of autonomy is just instrumental. It agrees that it is better to have ten options than to have just one, but says the only reason is the effect on other choices in the future. Someone with ten options now can practice choosing among them, and so become better at securing good options in the future. Without autonomy he would gain no experience in choosing, and suffer bad consequences later.

This point is serious, and important for any practical discussion of autonomy. But I do not think it solves the theoretical problem, or accounts for all the value of autonomy. For one thing, it does not capture the feeling many of us have that autonomy *is* intrinsically good, and that a life without free choice is poorer for that very reason. What is more, by assuming that there *will* be choices in the future, it does not tell against the most serious possible losses of autonomy. We can imagine a world—it might be brave and new—where people never make choices, but are always constrained into what they or some authority thinks best. In this world present restrictions on choice never have bad effects on future choice, yet this hardly makes them less objectionable. Finally, the view does not handle all the actual cases. To violate a citizen's autonomy the state need not try forcing her into one best activity. Instead of removing all but one of ten options it may remove just the bottom two, on the ground that they are intrinsically bad. This more limited coercion—forbidding the worst rather than requiring the best—still violates classical liberalism, and still (somewhat) reduces autonomy. Yet it may have no ill effects. The agent can still practice choosing among her eight remaining options, and, if she is protected from choosing the worst options, may receive a net benefit.

An opposing view retains the idea that autonomy is intrinsically good, but denies that this needs any justification. Moral foundations, it says, cannot be given for all things, especially not for all intrinsic goods. These goods come on a list—an “objective list,” some call it⁵—and if autonomy appears on the list there may be no deeper reason than that this is intuitively appealing. Any morality contains some ultimate commitments, some claims for which no deeper grounding is possible. That autonomy is intrinsically good should for us be one such commitment.

This view is not inane, and it may represent a fallback position. But it would be defeatist to adopt it from the start. We are challenged to explain why autonomy is good, and the challenge is not obviously inapposite. It asks whether autonomy can be connected with the other items on our list of intrinsic goods, and shown, like them, to reflect deeper values. If this was possible it would enrich and substantiate the claim that free choice has intrinsic value. How might it be done?

One suggestion derives from Mill's claim that an autonomous agent has “a character,” while one who is not autonomous has “no character, no more than a steam-engine has a character.” Now, if “character” means just “autonomously developed character,” Mill's remark is unhelpful. But there is another possibility. A person

⁵ Derek Parfit, *Reason and Persons* (Oxford: Clarendon Press, 1984), pp. 499–500.

has a character, we may say, when her choices are unified through time, and reflect some enduring traits and values. She has a character especially when her choices are unified in some distinctive way, so the traits revealed are special to her. And the suggestion then is this. A person who has autonomy can use it to develop her own values, and to express them consistently in action. But someone with no opportunity for choice cannot develop the same way. His acts can only reflect his momentary concerns, without the integration through time that makes for genuine personality.

These points are again serious, and important for a practical discussion of autonomy. But they still treat autonomy as just instrumentally good. The expansion of choice contributes not to better individual choices, but to a more desirable pattern of choices in one's life as a whole. It is instrumental at a higher level, but still instrumental. And for this reason, the suggestion does not value autonomy everywhere it should. Someone with free choice may use it to develop an integrated character, but then again he may not. He may freely flit from one passion to another. If he does, many of us will still think his autonomy a good, but the suggestion cannot explain why. By tying autonomy to a further good it only values autonomy when it results in that further good.

A more promising idea occurs in the second passage from Mill, about "perception, judgment, [and] discriminative feeling." Someone with many life options can deliberate about them, and in so doing exercise his rational powers. He can weigh the merits and defects of his various alternatives, and arrive at a reasoned conclusion about them. The presence of different possibilities calls for reflection upon them, and this reflection is a good. It exercises a high rational capacity, the capacity for discriminative valuing.

The defect in this second suggestion is more subtle. As the suggestion describes it, deliberation is an intellectual activity, concerned with assessing the merits of possible actions. And as such it does not require actually available options. Someone with no free choice—a slave, at the extreme—can still evaluate possibilities. He can ask what career would be best for him if he were choosing careers, and in so doing arrive, in principle at least, at the same intellectual grasp as someone with many careers open. Of course, a slave is unlikely to attain this grasp. People rarely deliberate about options they cannot choose, and this supplies a further instrumental argument for autonomy. By giving its citizens varied options a free society encourages the use of their deliberative powers. It encourages a reasoning that would have no practical point if options were closed. But the argument is still instrumental and does not give us what we want, which is deeper values present in the act of free choice itself.

II. Agency

To locate these values I suggest that we consider an ideal of agency. If one person chooses action *a* from ten options while another has only action *a* available, it may be true of each that she has made *a* the case, and is in that sense responsible for it.

But there is an important difference between them. The first or autonomous agent has also made certain alternatives to *a* not the case; if her options included *b*, *c*, and *d*, she is responsible for not-*b*, not-*c*, and not-*d*. The second person did not have this further effect. Since *b*, *c*, and *d* were not possible for her, the responsibility for their nonrealization rests with nature, with whoever constrained her—in any case, not with her. It is not in her person that those truths about her action originate. The ideal of agency is one of causal efficacy, of making a causal impact on the world and determining facts about it. And the autonomous agent, just in virtue of her autonomy, more fully realizes this ideal. When she chooses among options she has two effects: realizing some options and blocking others, and this gives her a larger efficacy than someone whose only effect is the first.

To illustrate, imagine that the first person chooses a career as a teacher from ten available careers, including lawyer, politician, and accountant, while the second becomes a teacher because that is her only option. If we ask why these people are teachers our answer may at one level be the same. But if we ask why they are not lawyers, politicians, or accountants, the answers will be different. In the first case the explanation will point to something inside the agent: to the fact that she chose not to be a lawyer, politicians, or accountant. But in the second case it will look outside the agent, to her society or to the people who limited her choice. As the difference in these explanations reveals, the first person is responsible for more facts about her life, and thus is more expansively an agent. To succeed as an agent is to make a difference in what the world does and does not contain, and this is more possible with more numerous options. By letting people decide what they do not do as well as do, autonomy makes them more widely efficacious.

This argument provides a sketch of a justification for autonomy, but to elaborate it we must specify our ideal of agency. To be plausible, this ideal must be one of *intentional* agency, of achieving goals one *intended* to achieve in advance. It is not any causal efficacy that has value, but efficacy that expresses some aim in the mind. And, because of this, autonomy has internal as well as external conditions. It is not sufficient for autonomous action that a person have many options open. He must, most obviously, know about the options, or he cannot intend their nonrealization. And he must also make in the fullest sense a choice among them. By this I mean a choice that is for one option *in preference to others*, so his rejection of the others appears in his mind. This does not always occur in intentional action. Someone who is driven by obsession may know that alternatives are available, but his moving intention does not reflect this. It goes blindly for *b*, if *b* is what he does, without preferring *b* to other options. (A strong claim is that the obsessed person intends only *b*, without rejecting anything. A weaker claim is that he rejects only the vague alternative not-*b*. Either way there is not the rejection of individually discriminated alternatives that on my view makes for autonomy.) The same is true of someone who is weak-willed. If a weak-willed agent acted on his best judgment that *a*, he would prefer *a* to other actions. But he does not succumb to temptation in preference to anything; he just succumbs. If autonomy achieves many intentions

it requires what I call choice in the fullest sense: a simultaneous realization of some possibilities and rejection of others, so one's knowledge of the others appears in and through what one wills.⁶

There is an objection to this argument that must be considered. I have proposed valuing autonomy under the heading of agency, as something that increases the intentions a person achieves, and makes her more expansively an agent. This may seem to presuppose that there is one correct individuation of a person's intentions, and that, as so individuated, her intentions can be precisely counted. Is this not absurd? Is it not absurd to say that someone has achieved 728 intentions, as opposed to 727? I agree that this is absurd, but it is not required for the argument. It is not in general required, for talk of more and less, that one can say exactly how much more or how much less. We talk of one person's having more freedom, or having more options, without believing that there is any cardinal measure of freedom. And we can take the same line with intentions. We can say that one agent has achieved more intentions than another without believing it possible even in principle to say exactly how many she had or carried out.

Autonomy, then, increases the goals a person achieves, and thereby increases her agency. But it does this even more if her choice follows deliberation. (I now take on Mill's point about "perception" and "judgment.") An agent who deliberates about her options may discover that *a* has the most of some desirable property *F*, that *b* has defect *G*, and so on. When she chooses, then, she intends not only *a* but the-option-with-the-most-*F*; and alongside *b* she rejects the-option-with-*G*. Her deliberative knowledge, if it guides her choice, gives her more intentions in and around her options than if she picked blindly among them. So it increases even further her agency. And there is another effect. If she has deliberated she will choose *a* and reject *b* as means to a single goal, perhaps getting-the-most-*F*-without-*G*. Her various intentions will converge on one aim, and by so doing form a hierarchical means-end structure that realizes sophisticated rationality. This will especially be so if her deliberations are part of some larger project. Then she will intend the-most-*F*-without-*G* as a means to some further goal, which perhaps serves further goals. Her intentions will be arranged thus, and she will achieve, not only many goals, but goals in a rational structure. If the exercise of rationality is a good, she will have it present in the intentions her choice carries out. What we have here is not just autonomy, but *deliberated autonomy*: free choice from a wide range of options that follows correct reasoning about them, and expresses its

⁶ Among the different conceptions of autonomy (see note 3), we may distinguish between external and internal conceptions. On an external conception autonomy means freedom from compulsion by external forces, whether based in other people or in nature. On an internal conception it means self-mastery, or freedom from compulsion by internal forces such as obsession or uncontrollable desire. My conception of autonomy as a choice from many options is primarily external. But, as the last paragraph reveals, it also has internal components. I do not claim that my conception captures everything that has been intended by adherents of the internal view, but it does capture some of it.

results. This autonomy exercises our rational powers, as Mill's argument notes. But it does so in a more than intellectual way. The elements it organizes are not beliefs, which are available to a slave, but successful intentions about the world. And because of this it presupposes the simpler autonomy that consists in any choice among options. The foundation of deliberated autonomy is a set of acceptances and rejections that converges on one goal, and this is only possible for someone with many options to select or reject. There can only be a rational hierarchy around her choice if she has a real choice of options to begin with.

We have, then, two cases: simple autonomy, or any free choice among options, and deliberated autonomy, which follows careful reasoning about them. Of these the second is a greater good. It involves a greater realization of agency, and combines agency with the distinct good of rationality. But in both cases the basic value of choice is the same. An autonomous agent, as autonomous, achieves more ends, whether organized or not, than someone with no choice to make.

In reducing autonomy to agency I do not identify the notions, for autonomy is just one part of agency. We call a person autonomous when he can make choices about his own life; it is something beyond this to make choices about other people's lives, or about the material world. A person has autonomy if he can decide whether or not he eats a banana; he has something more—call it external power—if he can decide whether someone else eats a banana, or whether the banana falls from the tree. A morality that values autonomy for realizing agency should also value external power, and seek to increase it for all humans. But here there is a difficulty. Because it appears within lives, autonomy can be enjoyed by all humans, without one's hindering another's. But power is essentially competitive. My having more inevitably means your having less. The politics of external power are therefore more complex than those of autonomy, even though the basic values in each case are the same.

My account of autonomy may illuminate a part of Marx's value theory. The central values in Marx's morality emerge from his philosophical anthropology, from his view of humans as essentially productive and cooperative. Why then the emphasis in Marx's writing about the future on "the free development of each," and the "free play" of humans' "physical and mental powers"?⁷ What would be wrong for Marx with forcing people into productive, cooperative activity? Marx happens to think this forcing will be unnecessary; under communism people will have work as "life's prime want." But this does not resolve the puzzle. Marx clearly thinks the freedom of communist labor is not just another fact about it, but something that adds to its value. And what justifies this view? The idea of autonomy as agency may help. If autonomy is a matter of acting in the world, of determining facts

⁷ Karl Marx and Friedrich Engels, *The Communist Manifesto*, in D. McLellan, ed., *Karl Marx: Selected Writings* (Oxford: Oxford University Press, 1977), p. 236; and Marx, *Capital* (New York: Vintage, 1977), vol. 1, p. 284.

about it, it may fit a theory where productivity is centrally valuable. Autonomous agents may achieve in their own lives what workers achieve in objects by productive labor upon them.

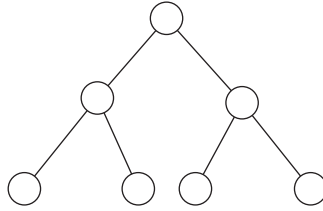
III. Knowledge and Quality

I have tried to explain the worth of autonomy by connecting it to deeper values, in this case the values of agency. But the argument can be taken further by considering knowledge, another item on many lists of intrinsic goods. To have knowledge is to stand in a certain relation to the world, the relation where one's beliefs match what the world contains. And those who value knowledge think it better to stand in this relation on more occasions. If knowledge consists in true beliefs, it is better to have more true beliefs. If someone acquires a new piece of knowledge without losing any old ones this is good; if an injury destroys knowledge this is bad. Knowledge involves a relation of correspondence between one's beliefs and the world, and, on a plausible view, the more items of knowledge one has the better.

Agency as I have described it is the obverse of knowledge. It too involves a relation of correspondence between one's mind and the world, but the direction of correspondence is different. Instead of one's beliefs conforming to the world, the world comes to match one's aims, through successful intentional action.⁸ And again there is a value in number. Just as it is better to have more items of knowledge, so it is better to have more intentions one has achieved. Behind agency, in other words, we can see a more general value of relation-to-the-world, or perhaps life-in-the-world. It has a theoretical component, consisting in knowledge, and a practical component, namely agency. And each of these has a self-referential sub-component. On the theoretical side there is self-knowledge, or knowing facts about oneself, and on the practical side self-determination, or determining facts about oneself. If autonomy increases self-determination it increases agency, which in turn increases a general relation-to-the-world. So autonomy is no isolated value, but one manifestation of a deep ideal of connection to reality, both within oneself and outside it.

If the parallel with knowledge deepens our account of autonomy, it also has another effect. Those who value knowledge do not think it just a matter of having many true beliefs. Obviously, some kinds of knowledge are more worth having than others. It is better to know a fundamental law of the universe than to know the number of redheads in North Bay, Ontario, and better to understand the workings of a friend's personality than to know the exact length of his forearm. A view

⁸ On most views knowledge is more than just true belief, it is *justified* true belief. This can be mirrored in the account of agency. We can value only those successful intentions that were preceded by a justified belief that they would be successful.

FIGURE 8.1 *Structured Goals*

that values knowledge intrinsically needs a test for discriminating among items of knowledge, for saying which have more and less worth. It needs a standard of quality, and here a plausible standard is explanatory: the best knowledge is the most explanatory, explaining the most other knowledge and accounting for its truth. Ideal knowledge, we may say, comes arranged in structures (see Figure 8.1) with general principles that explain derived particulars. The most valuable items of knowledge are those at the top of these structures. A scientific law explains many individual phenomena, a general truth about someone's character much of his particular behavior. By explaining many truths below them, these items of knowledge have great explanatory importance, and great intrinsic worth.

An account of agency needs a parallel standard of quality, for plainly some achievements are more valuable than others. It is better to complete a year-long program of self-improvement than to successfully tie a shoelace; and better to reform one's whole society than to light a cigarette. Again a plausible test of quality is structural: the best achievements organize large numbers of others, by having those others as means to them. Like beliefs, intentions can come in hierarchies, with those lower down pursued as means to those above. And an attractive view makes the best intentions the ones atop such hierarchies. By organizing many others they have the most practical importance, and their achievement has the most intrinsic worth.

I have hinted at this view in the account of deliberated autonomy. One virtue of the choice that follows deliberation is that it organizes many intentions in one structure, and exercises sophisticated rationality. But there is another implication. If autonomy realizes agency, and agency admits differing values, then some autonomous choices—those of organizing goals—have more intrinsic worth. It may matter a little that one ties a shoelace autonomously, but it matters a great deal that one chooses a year-long goal autonomously. And the most important autonomy of all concerns whole plans of life, or whole sets of values. By organizing all a person's activities, these goals sit atop immense hierarchies, and their autonomous choice has immense intrinsic value.

This is a point of contact with Humboldt and Mill, whose chief concern is always the autonomous choice of whole forms of life. They care most about freedom to fix the general shape of one's life, or one's general moral values. Their reason here is partly instrumental. If free choice has extrinsic benefits, these are doubtless

greatest when the choice is most far-reaching. But I think they also have an intrinsic concern. If autonomy is valuable, our intuitions suggest, it is not uniformly valuable. Some free choices have more intrinsic worth, as choices, than others, and by exploiting the parallel with knowledge we can explain which they are. It is most valuable to choose goals that organize and encompass many others subordinate to them in a means-end hierarchy.

There is a question of terminology. I have said that the autonomous choice of organizing goals is more valuable than the autonomous choice of trivial goals. Is this because the first choices are more autonomous? I have no fixed view about this. It concerns the use of the English word “autonomous,” and given its many ambiguities (see note 3) there may be no clear answer. So I will say simply that the free choice of organizing goals has more value—of whatever kind—than the free choice of specific goals. Consider the person who flits freely from one passion to another. His choices, I have suggested, are less valuable than those of someone who freely chooses a unified life plan. But I cannot decide whether this is because his choices are less autonomous, or because they are lacking in some other, and separate, value.

There may also be contact with Humboldt’s and Mill’s interest in a “variety of situations” and “experiments of living.” This interest is again partly instrumental. Humboldt and Mill believe that different life forms are suited to different people, and that detailed information is needed to choose wisely among them. But they may also intend something intrinsic. As we have seen, autonomy requires knowledge of available options. But it seems fullest when this knowledge is greatest, when the options are grasped not just vaguely but in articulated detail. Again the value of hierarchy can explain why. Someone who lives in a lively, diverse society can choose among life forms he knows at first hand, and, when he chooses, he can intend not just loose generalities but connected series of acts. Within a large hierarchy that unites choices and rejections he can have many smaller hierarchies representing all the worked-out options he declines.

The value of hierarchy also helps answer another objection to my proposal. If autonomy is a matter of achieving additional ends, it may be said that we should, when performing some trivial act like lifting a fork, think of all the other things we are not doing and consciously reject them. We should think that we are not lifting our knife, or using our other hand, and intend that all these things be so. Surely this is absurd. Surely these extra rejections give our action no greater worth.

The reference to consciousness in this objection is a red herring. Just as we can have knowledge that we are not now considering, so we can achieve ends that are not present to the mind. As athletes, dancers, and car drivers remind us, we can act intentionally, and act on intentions, without having any goals in consciousness. And, even apart from this, the objection fails because it ignores the dimension of quality. The alternatives we reject when we lift a fork are all highly particular, and as such of minimal value. Whether it is autonomous or not, manipulating cutlery is a trivial exercise of agency. And, because of this, if

troubling to do it autonomously distracts us from greater goods it will be on balance wrong. To seek greater choice here is not to seek what is best. There is always a parallel with knowledge. Knowing the number of redheads in North Bay is so trivial that someone who actually seeks this knowledge will be wasting her time. When there are so many better things to strive for, attention to that will be counterproductive.

There is a stronger response possible to the objection. Some may think the number of redheads in North Bay so trivial a truth that knowing it adds nothing at all to a life. It is not that this knowledge has some minimal value that is outweighed by instrumental disvalue; it has no value at all. I do not incline to this view, but it is certainly available. And to capture it we could introduce a threshold on the scale of quality, saying that knowledge below this threshold has no intrinsic value. Without some minimal explanatory power, we may say, its presence counts for nothing. A similar view is possible for agency. It holds that achievements below a threshold of organizing importance, and *a fortiori* autonomous achievements below the threshold, have no value at all, and are never intrinsically worth seeking. Unless an end has some others subordinate in a means-end hierarchy, its achievement, whether autonomous or not, has no intrinsic worth.

Whichever response we choose, our central thesis remains. It holds that there is intrinsic value in the autonomous choice of important goals, ones that structure a day, year, or life. And the value of these choices can be captured under the heading of agency. Someone who chooses, in the fullest sense, among forms of life determines not only that she lives in one way, but that she does not live in others. Her autonomy makes her responsible, not just for the positive fact that her life has one shape, but for the negative facts that it does not have various other shapes. She is more efficacious in the world, specifically in the part containing her life, than someone who has only one form of life available. Her autonomy is especially valuable if it follows deliberation, and chooses among options that are fully understood. Then her intentions are hierarchically arranged, and exercise sophisticated rationality. But the basic value remains that of agency: to be autonomous, to choose among life options, is to determine more of what is true than if one is constrained. To have ten options rather than just the best among them is to be able to say no as well as yes. It is to be able to say no nine times, and to be responsible for the fact that no was said.

IV. Autonomy and Liberalism

I want to close by discussing the political implications of this argument. In introducing the view that autonomy is intrinsically good I said that it was part of the classical case for liberalism. If we believe, as Humboldt and Mill believed, that free choice is intrinsically good, we have one reason to place limits on the

legitimate power of the state—specifically, to forbid the state to interfere with citizens' liberty except to prevent greater losses of liberty for others.⁹ But how strong a reason is this? How much of political liberalism does a view about autonomy support? And, in particular, how much does it support if it is grounded in the way I propose? Humboldt and Mill were not just liberals, they were unqualified liberals. They held that, in the world as it is, the state must *never* interfere in the private lives of citizens. How much of this stand follows from a perfectionist valuing of autonomy?

Even if we consider autonomy alone, the stand cannot follow entirely. Our view values free choice as an intrinsic good, but it could be that restricting a person's choice now did more to increase his own choice later, or to increase the choices of other people. By denying him present options the state could ensure that he had more options in the future, or a greater capacity to choose autonomously among them. So violating some liberty now could increase his total autonomy through time. Analogously, denying one person options—for instance by enslaving him—could in principle increase the options available to others, and thus perhaps increase aggregate social autonomy. These possibilities, especially the last, may not often be realized. But they are one bar to a simple argument from autonomy to unqualified liberalism. Even if free choice has intrinsic value, some restrictions on it may be justified by the need to preserve and promote free choice itself.

There is, however, a more serious worry. I have argued that a plausible value-theory should treat autonomy as one intrinsic good, as one item on a list of perfections. But it cannot treat autonomy as the only intrinsic good. To be plausible it must acknowledge that there are other goods, and that, as a result, some autonomous states and activities are better than others. Autonomy is good, yes, but autonomous creativity is better than autonomous idleness, and autonomous knowledge better than autonomous ignorance. But this may seem to reopen the door to illiberality. If autonomy is just one good among others, may sufficient increases in those others not outweigh any loss in autonomy? And may they not thus license much interference with the self-regarding? The worry here is not that valuing autonomy gives no support at all to liberalism. If free choice is intrinsically good, every limitation of it carries some moral cost. But must this cost

⁹ This is a traditional understanding of liberalism, as a doctrine about liberty, and differs from the more recent understandings of John Rawls, Ronald Dworkin, and Bruce Ackerman; see Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), pp. 31, 446–52; Dworkin, "Liberalism," in Stuart Hampshire, ed., *Public and Private Morality* (Cambridge: Cambridge University Press, 1978), pp. 113–43; and Ackerman, *Social Justice in the Liberal State* (New Haven: Yale University Press, 1980). According to these writers a liberal state must not only refrain from forcing people into some ideal of the good life, it also must not do anything to promote that life above others. It must be neutral about theories of intrinsic value, including the theory that autonomy is intrinsically good. Needless to say, the arguments of this paper do nothing to support "liberalism" in this more restrictive, nontraditional form.

always be decisive? If there are goods other than autonomy, may sufficient gains in them not compensate for any loss in autonomy?

I can think of two ways to resist this conclusion. One is to treat autonomy as a good among others, but give it vastly greater relative weight. At the extreme, autonomy can be made lexicographically prior to all other goods, so that increases in it are infinitely more important than increases in anything else. More weakly, autonomy can be given just much more finite weight. Either way, the connection with strong liberalism is retained. If autonomy counts vastly more than other goods, increases in those goods will almost never justify a diminution of autonomy. If free choice has tremendous relative weight, its preservation will almost always have moral priority.

Unfortunately, this view has certain counterintuitive consequences. Imagine that someone has chosen an act fairly autonomously, from a fairly wide range of options. If she had a few more options her act would be slightly more autonomous, and slightly better. But is this improvement more important than any other improvement in her act's quality? Our intuitions care a great deal about the difference between no autonomy and substantial autonomy, but much less about the difference between substantial and enormous autonomy. Yet if autonomy has tremendous relative weight we must prefer increasing slightly the freedom of already free individuals to making any other improvement in the worth of what they do. The view also goes wrong in some of its comparative judgments. I have said that autonomous creativity is better than autonomous idleness, and most of us would say it is considerably better. But if autonomy has tremendous relative weight this is not so. Suppose that a creative act earns 1,000 units of value for being autonomous and 10 for being creative. If idleness also earns 1,000 units for being autonomous, there is too small a percentage difference between them.

If we are impressed by these difficulties we may shift to a slightly different view. It treats autonomy not as a good among others, but as a condition of goods. It holds that if an act is minimally autonomous, its other virtues can be assessed in a standard way. But if the act is not autonomous—if it falls below some threshold of autonomy—it has no value at all. Without some free choice an act's other qualities do not enter the moral reckoning. This view is not as strongly liberal as its predecessor, for it allows some losses in autonomy to be justified by gains in other goods. Mild restrictions on choice that leave an act above the threshold of autonomy may be permitted, as may more serious interferences that improve other aspects of other activities. It may be permitted to make checkers nonautonomous if that brings about sufficient improvements in people's autonomous playing of chess. But the view does exclude the most obvious illiberality. If autonomy is a condition of goods, serious restrictions of it cannot be justified by other improvements in the very same activity. If a law requiring chess makes chess unfree, it cannot be justified by other improvements in people's chess technique.

Either of these views, then, retains the link between autonomy and a fairly strong liberalism. But neither is available if the main argument of this paper has succeeded. That argument set out to explain the value of autonomy by connecting it to other values, by showing that it realizes the goods of agency and, more deeply, relation to the world in one special area. Free choice, it held, is no isolated good, but follows with other goods from more fundamental values. But this argument, if successful, precludes any special treatment of autonomy. If autonomy is just one manifestation of deeper values, why should it count more than other manifestations like external agency and knowledge? And, if they are not conditions of goods, how can it be a condition of goods?

This, then, is our situation. To move from valuing autonomy to a strong liberalism we must treat autonomy differently from other goods. But if we explain the value of autonomy by connecting it to other goods, this is no longer possible. Autonomy is just one value among others, with no special weight and no special standing.

How we respond to this situation depends on our attitude to strong liberalism. Those who do not find strong liberalism plausible will have no difficulty. They will agree that autonomy is one intrinsic good, and that as a result there is always some moral cost in a restriction of liberty. But they will insist that this cost need not always be decisive. If interfering with self-regarding action does enough to promote goods other than autonomy, it may be on balance justified.

The more difficult position is that of those who do find strong liberalism plausible. They could respond by giving up the attempt to explain the value of autonomy, and instead treat choice as an ultimate good with no deeper grounding. This is the fallback position I mentioned earlier. But to adopt it would in my view seriously weaken the strong liberals' position. It would deprive them of an argument that enriches and substantiates the claim that autonomy is intrinsically good. And they have in any case another response available. They can, while using the intrinsic value of autonomy as one support for liberalism, appeal to other supports for the same conclusion. They can, most obviously, appeal to the various instrumental arguments for autonomy mentioned at the beginning of this chapter. And they may also appeal to some different arguments about utility or perhaps even libertarian rights. An ideal of the person as autonomous and self-directing accompanies the unqualified liberalism of Humboldt and Mill, and is part of its rationale. But it need not be all of its rationale. Even if autonomy alone does not place an absolute limit on the legitimate powers of the state, it may be that a collection of considerations, of which it is just one, does support this strongly liberal conclusion.

It is not my purpose to decide between these responses, or to reach a final assessment of strong liberalism. I have tried merely to substantiate the view that free choice is intrinsically good by connecting choice to deeper perfectionist values. And on any reading this view gives some support to political liberalism. If

autonomy is intrinsically good, any restriction of choice has some morally undesirable effects, and there is always some reason to resist state interference with the self-regarding.¹⁰

Postscript

There is a shorter presentation of the ideas in this chapter in chapter 11 of my book *Perfectionism*.

¹⁰ The central challenge of this paper was first put to me by G. A. Cohen; my response to it emerged in conversation with David Copp. For further comments I thank Derek Allen, G. A. Cohen (again), Gerald Dworkin, Dennis McKerlie, Donald Regan, Wayne Sumner, David Zimmerman, and an acute audience at the University of Lethbridge.

Desert: Individualistic and Holistic

This chapter is a response to a recent argument of Samuel Scheffler's. Scheffler is defending John Rawls's view that while the concept of desert may play a foundational role in the theory of retributive justice that underlies the criminal law, it cannot do the same in the theory of distributive justice that assesses economic outcomes.¹ The concept of desert, Scheffler argues, is individualistic, in that what a person deserves depends only on facts about him. This suits it for the context of retributive justice, which is likewise individualistic. But distributive justice is essentially holistic, since the justice of a given person's share of resources turns in several ways on facts about other people. This mismatch between the individualism of desert and the holism of distributive justice, Scheffler concludes, best justifies Rawls's rejection of foundational distributive desert.²

I do not agree that the concept of desert is essentially individualistic. Like many, I believe that desert has not only individualistic but also holistic aspects, or, as some would say, not only noncomparative but also comparative aspects. But I propose to explore how these aspects relate in the specific contexts of moral, retributive, and distributive desert. I will argue that the most plausible theory of economic or distributive desert is in two ways more holistic than the most plausible theories of moral and retributive desert. This conclusion in a way mirrors Scheffler's, since it emphasizes the holism of distributive justice, but it also undercuts it. If economic desert is holistic in the ways I will describe, it does not ignore but precisely acknowledges the holism of the distributive realm.

¹ J. Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), pp. 314–15.

² S. Scheffler, "Justice and Desert in Liberal Theory," *California Law Review* 88 (2000): 965–90, and "Rawls and Utilitarianism," in his *Boundaries and Allegiances* (Oxford: Oxford University Press, 2001), pp. 149–72.

I. Moral Desert

The theory of moral desert holds that the morally virtuous deserve pleasure and the morally vicious deserve to suffer. It combines individualistic and holistic aspects in an especially straightforward way and is therefore a model for retributive and economic desert. My discussion of its two aspects will follow general lines laid down by Joel Feinberg and Shelly Kagan,³ but will differ from them on several points of detail.

As I understand them, claims about desert assign intrinsic value to combinations of states as combinations, or in addition to the values their parts have on their own. According to an individualistic principle of moral desert, the relevant combinations are always of states within the life of a single individual. If a virtuous person enjoys pleasure, then his virtue is good, his pleasure is good, and there is a further good, which I will call a desert-good, in the fact that these two goods occur in the same life. If a vicious person suffers pain, there is a similar desert-good in the fact that these two evils occur in the same life. But if a virtuous person suffers pain or a vicious one enjoys pleasure, he is getting the opposite of what he deserves and his combination of states is a desert-evil.

In addition, I will assume this individualistic principle has what I call an optimality structure: it says that for any degree of virtue or vice there is a quantity of pleasure or pain that it is best in desert terms for a person to have, with greater and lesser quantities having less desert-value and in extreme cases being evil. I will also assume that the principle satisfies a proportionality condition, according to which the best division of a fixed quantity of pleasure or pain among individuals is always proportioned to their degrees of virtue or vice, so a person who is twice as virtuous enjoys twice as much pleasure. This ideal of proportionality has been prominent in the classical literature on desert and can be included in an individualistic principle by means of this condition. The resulting principle is represented in Figure 9.1, where each curve shows how the desert-value of a person's pleasure or pain is a function of its quantity, given a fixed degree of virtue or vice in his character. The optimal quantities of pleasure and pain are represented by the peaks on the curves, and two points about these peaks' locations deserve comment. First, the peak for a person who is twice as virtuous or vicious is always twice as far to the right or left of the origin, so his optimal pleasure or pain is twice as great. This is an obvious implication of proportionality, but the peak for a person who is twice as virtuous or vicious is also always twice as high, so his having that optimal pleasure or pain has twice as much desert-value. This, too, follows from proportionality, and is intuitively attractive. If we had the choice between

³ J. Feinberg, "Noncomparative Justice," in his *Rights, Justice, and the Bounds of Liberty* (Princeton, N.J.: Princeton University Press, 1980), pp. 265–306; S. Kagan, "Equality and Desert," in L. P. Pojman and O. McLeod, eds., *What Do We Deserve? A Reader on Justice and Desert* (New York: Oxford University Press, 1999), pp. 298–314.

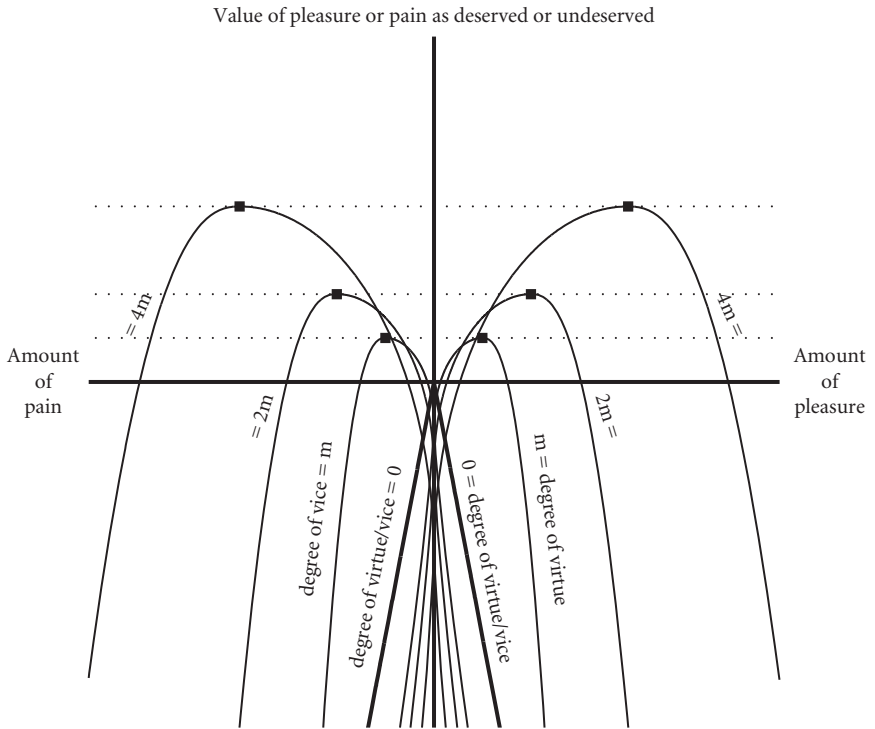


FIGURE 9.1 *Individualistic Moral Desert*

giving the optimal reward to a saint or to someone only slightly virtuous, surely it would be better in desert terms to give it to the saint; if we could inflict the pain he deserves on someone slightly vicious or on Hitler, it would be better to inflict it on Hitler.

This principle has a further attractive feature. In a simpler graph the curves would all pass through the origin, implying that a virtuous or vicious person's experiencing neither pleasure nor pain has neutral value. But it is more plausible to hold that the absence of any reward for the virtuous or punishment for the vicious is positively evil, and curves that express this idea will cut the vertical axis below the origin. They do so in Figure 9.1, and in a way that has further attractive implications. The curves for people who are more virtuous or vicious cut the axis further below the origin, so failing to reward or punish greater virtue or vice is a greater desert-evil. These curves also cut the horizontal axis further to the right or left, so the zero-value quantity of pleasure or pain for greater virtue or vice is likewise greater. A quantity of pain that is positively good for a mildly vicious person can be evil, because so utterly inadequate, for a horribly vicious one.

The principle represented in Figure 9.1 is individualistic because it assigns value only to combinations of states within individuals' lives. It has implications for distribution, and in particular prefers distributions that are proportioned to people's degrees of virtue or vice. But this preference does not result from its valuing proportional distributions as distributions; it is a side effect of the specific way it values combinations of states within single lives.

A holistic principle of moral desert does value distributions as distributions. More specifically, it values patterns of distribution in which people's shares of pleasure and pain match their degrees of virtue and vice, most commonly by being proportioned to them. If those who are twice as virtuous enjoy twice as much pleasure, that is best as a pattern of distribution; if they enjoy either more or less than twice as much, that is less good and, as the disproportion increases, can become evil. This holistic principle is therefore more complex than the individualistic one: it values not combinations of states within individual lives, but combinations of those combinations that instantiate a pattern of proportionality, one where the ratio of people's pleasure or pain equals the ratio of their virtue or vice.

The best representation of this principle is on another graph, one whose horizontal axis measures the pleasure or pain of person *A* and whose vertical axis measures that of person *B* (Fig. 9.2). Given a fixed proportion between these persons' virtue or vice—and in this graph *A* is assumed to be twice as virtuous as *B*—we can draw a ray out from the origin representing proportional distributions, ones that give the first person exactly the right multiple of pleasure or pain of that given the second. Since in this graph *A* and *B* are both virtuous, this ray is in the top right quadrant. If they were both vicious, it would be in the bottom left; if one was virtuous and the other vicious, it would be in the top left or bottom right. But for any two people there is some ray that represents perfectly proportioned distributions of pleasure or pain between them. Call this ray the proportionality ray and give all the points on it the highest value as distributions (say, five units). Now consider the contrary ray, the one going out from the origin in the opposite direction. It represents what I will call contra proportional distributions, which in Figure 9.2 involve *A*'s suffering twice as much pain as *B*. These are the holistically worst distributions and have the lowest value (say, negative ten units). This treatment of contraproportionality has the following rationale. Someone who intentionally produces a contraproportional distribution is responding in exactly the opposite of the appropriate way to the bases of moral desert. He has what we can call a satanic attitude to desert, treating someone's virtue as a reason not to give him pleasure but to give him pain and making those who are twice as virtuous not twice as well off but twice as badly off. So the distribution he produces should be holistically worst. Finally, for any distribution of pleasure or pain other than a proportional or contraproportional one, draw the ray out from the origin that contains the point representing it, measure the angular distance that ray is rotated around from proportionality to contraproportionality, and use that distance to give the point a holistic value between five and negative ten. In Figure 9.2 there are

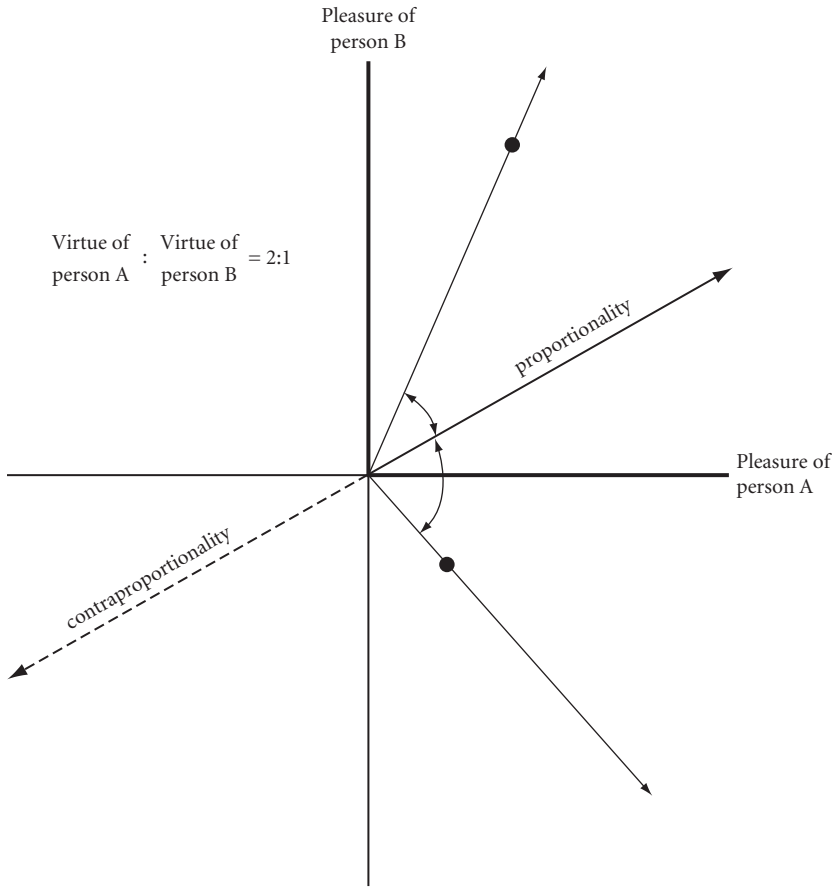


FIGURE 9.2 *Holistic Moral Desert*

two points represented by dots, one in the top right representing more pleasure for less virtuous *B* than for more virtuous *A*, the other in the bottom right representing pleasure for *A* and pain for *B*. Since in this case the point in the bottom right is rotated further around—the arc from its ray to the proportionality ray is longer—its holistic value is lower.

Given these accounts of the individualistic and holistic principles, does the best theory of moral desert contain only one or only the other or both? The principles agree that the best division of a fixed quantity of pleasure or pain among people is proportioned to their degrees of virtue or vice, and therefore overlap at many points. But each has gaps the other can fill, so the best theory of moral desert combines them both.

These gaps are most evident in the holistic principle. Since it values only a pattern of distribution across persons, it does not apply to situations involving only a single person. If all we know is that one person with a certain degree of

virtue is enjoying a certain quantity of pleasure, the holistic principle says nothing at all about the degree of desert-value this situation contains. In addition, the principle is indifferent between very different instantiations of the same distributive pattern. If *A* and *B* are equally virtuous, it assigns the same value to the situation where they both enjoy one unit of pleasure and the situation where they both enjoy a thousand units. If we find these gaps unacceptable, we can supplement the holistic principle with an individualistic one. It precisely does find desert-value in situations involving a single person, and prefers the outcome where *A* and *B* enjoy a thousand units of pleasure if that places them closer to the peaks on their individual desert curves.

But the individualistic principle also has gaps. Imagine that *A* and *B* are equally virtuous and both enjoy less pleasure than they ideally deserve, so they are both to the left of their desert peak, but *B* enjoys considerably more pleasure than *A*. There is a disproportion in this situation, and the individualistic principle says it would be better if this were removed by increasing *A*'s pleasure to the level of *B*'s. But what if the disproportion is removed in the opposite way, by reducing *B*'s pleasure to *A*'s? The principle says this "leveling down" only makes the situation less good, by replacing a greater desert-good with a lesser one. But I think many of us will see it as in one respect an improvement. The holistic principle captures this view, saying the reduction in *B*'s pleasure removes a holistic evil of disproportion. Or imagine that *A* and *B* both enjoy more pleasure than they deserve, so both are to the right of their peak, but *B* again has more. The individualistic principle says it would only make things worse if *A* were raised to *B*'s level, but the holistic principle, again in my view more plausibly, says this leveling up would in one respect improve them. These claims are not uncontroversial. Some philosophers vigorously reject leveling down, calling the idea that some should be denied what they individually deserve because others are not receiving what they deserve "plainly unacceptable."⁴ But I think our beliefs about desert include a concern with patterns as patterns that is not completely captured even by an individualistic principle that satisfies the proportionality condition. Just as the best theory of moral desert must supplement a holistic principle with an individualistic one, so it must supplement an individualistic principle with one that values proportional distributions as such.

As I have described this theory, its individualistic and holistic components are mutually independent. Each can be formulated without reference to the other and each can therefore be accepted apart from the other. This is most obvious for the individualistic principle, since one can talk about what people individually deserve without mentioning any pattern of distribution across them. But it is also true of

⁴ P. Montague, "Comparative and Non-Comparative Justice," *Philosophical Quarterly* 30 (1980): 133; see also J. Hoffman, "A New Theory of Comparative and Noncomparative Justice," *Philosophical Studies* 70 (1993): 173–74, 181 n. 16.

the holistic principle, since one can value a pattern of proportionality between people's receipts and their virtue or vice without mentioning individual desert. At the same time, however, the principles overlap in their claims about the best division of a fixed quantity of happiness, which they both say is proportioned to virtue. This mix of relations exactly parallels those between the broadly egalitarian views Derek Parfit has distinguished as equality (in the strict sense) and priority.⁵ The equality view is a holistic one valuing equal patterns of distribution as patterns; the priority view is individualistic, assigning values to states of individuals in such a way that a unit gain for a person who is worse off always counts more than a similar gain to one who is better off. These views are again independent, since neither's formulation refers to the other, but they agree about the best division of a fixed quantity of happiness, which they both say is always equal. The two views can disagree about other cases, however, especially ones involving leveling down, which the equality view favors but the priority view does not.⁶

If the best theory of moral desert includes a holistic principle, then it is not true that this form of desert is purely individualistic. Scheffler may reply that, unlike the kinds of justice he is interested in, moral desert has no political relevance, since the values it defines are not ones governments are called on to promote. So let us turn to the forms of desert that are politically relevant, beginning with retributive desert.

II. Retributive Desert

By retributive desert I do not mean the view that the morally vicious deserve punishment in an afterlife, which I take to be part of moral desert. Instead, I mean the view that those who commit criminal offenses deserve specifically legal punishment. It differs in several respects from the moral theory.

First, the theory of moral desert usually has a whole-life form, holding that on the basis of the virtue or vice in their lives as a whole people deserve happiness or suffering in their lives as a whole.⁷ But legal retributivism is not a whole-life

⁵ D. Parfit, *Equality or Priority?* The Lindley Lecture (Lawrence, KS: University of Kansas Press, 1995).

⁶ The theory of moral desert I have described shares the same general structure as Kagan's but differs in four main ways. (1) Kagan's holistic principle is not independent of his individualistic principle but is formulated in terms of it (see "Comparative Desert;"). (2) Kagan does not think the individualistic and holistic principles need overlap about the best division of a fixed quantity of happiness. (3) His holistic principle does not value proportional distributions (see "Comparative Desert"). (4) Nor does his individualistic principle prefer proportional distributions. None of the individualistic principles he considers in "Equality and Desert" satisfies the proportionality condition, and the "bell motion" feature he introduces positively conflicts with proportionality.

⁷ This is emphasized in W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 56–57.

theory. It holds that those who perform specific criminal acts deserve specific punishments, in each case largely independently of their acts or happiness at other times. In addition, the property that grounds retributive desert, though related to moral virtue, is distinct from it. Forms of vice that involve only feelings, for example of envy or malice, but no intention to act merit no legal punishment. And even where a person does have a vicious intention, other factors, including ones beyond his control, can affect whether and how much punishment he deserves. If *A* and *B* both intend to commit murder but *A* is restrained by bystanders whereas *B* is not, *B* deserves punishment for a criminal act but *A* does not. If *A* and *B* both shoot at their intended victim but *A* misses whereas *B*'s bullet hits home, *B* deserves the greater punishment for murder and *A* the lesser one for attempted murder. In each case *A* is no less vicious than *B* but, by luck, is less criminally liable. Something analogous happens on the side of punishment. Though the punishment a criminal deserves is in general something that will cause him pain, it is specified in objective terms (for example, as so many years in prison) and is not adjusted up or down depending on exactly how unpleasant it is for him.

Second, legal retributivism does not mimic the whole structure of moral desert. The moral theory treats positive and negative desert symmetrically, finding the same desert-value in rewards for the virtuous and punishments for the vicious. But retributivism concerns only negative desert; it values only the punishment of legal offenders and does not mandate rewards for some contrary of criminal conduct. In addition, though retributivism finds it evil if the innocent are punished, it does not distinguish between their degrees of positive virtue. What matters and is contrary to desert is only that someone who did not commit a crime is punished as if he had. These differences mean that if retributivism contains individualistic and holistic principles, the graphs representing them will mirror only parts of the graphs for moral desert. The individualistic graph will parallel only the part of the moral graph to the left of the vertical axis, the part representing pain, and will eliminate the curves for people with positive desert, retaining only those for people who are negatively deserving or innocent, with the latter represented on the $v = 0$ curve. The holistic graph will define proportionality only in the bottom left quadrant, where both people suffer pain, and will take the axes bounding that quadrant to represent what innocent people deserve, which is no punishment at all.⁸

As so understood, both the individualistic and holistic principles are essential to a plausible retributive theory. This is most obvious for the individualistic prin-

⁸ Because it defines proportionality only in this one quadrant, this holistic principle avoids Kagan's various objections to the "ratio" or proportionality view of holistic desert (see "Comparative Desert"). The restriction to the one quadrant in effect denies what Kagan calls "optimism" about such desert. The holistic economic principles discussed later in this chapter likewise define proportionality only in one quadrant and therefore likewise avoid Kagan's objections.

principle. In fact, legal punishment may be the context where the optimality structure I have associated with individualistic desert is most intuitively compelling. Surely all retributivists hold that for every offense there is an ideally deserved punishment, and that punishments more severe than that are positively unjust, as capital punishment would be for car theft or even rape. Other features of the structure in Figure 9.1 are also attractive for retributive desert. Thus, it is plausible that individualistic retributivism satisfies a proportionality condition, so the optimal punishments for more serious offenses are more severe, and by an amount that matches their greater seriousness. It is also plausible that failing to punish an offense is not just not good but evil, and more evil for more serious offenses, and that a punishment that is good for a minor offense can be evil for a horrible one, as a six-month jail term would be for car theft on the one hand and murder on the other.⁹ Finally, Figure 9.1 has another feature that is attractive for individualistic retributivism. For any peak in the graph, the vertical line running down from it to the horizontal axis is shorter than the continuation of that line running from the axis down to the $v = 0$ curve. This implies that the desert-goodness of inflicting a given punishment on a person who ideally deserves it is always less than the desert-evil of inflicting that punishment on someone who is innocent, thus giving an axiological rationale for the common view that the legal system should be more concerned to avoid punishing the innocent than it is to punish every one of the guilty.¹⁰

But a complete retributive theory also requires a holistic principle, one concerned with the pattern of criminal punishments as a pattern. This principle is one aspect of the more general concern that the legal system treat all citizens evenhandedly or offer them “equal protection of the laws.” And it has at least two practical implications.

The first of these, which has been identified by David Dolinko, is to give some support to a system of fixed uniform sentences, as in the Federal Sentencing Guidelines currently in effect in the United States, as against one giving judges and juries wide discretion to determine sentences themselves.¹¹ The basis for this support is somewhat subtle. The holistic principle requires not only that like cases be treated alike, as Dolinko emphasizes, but also that unlike cases be treated in an appropriately unlike way. So if judges with wide discretion can recognize differences between criminal offenses that a scheme of uniform sentences

⁹ To say a six-month prison term for murder is evil is not to say it is worse than no punishment at all; on the contrary, no punishment for murder is worse than six months in prison. Instead, to say a six-month prison term for murder is evil is to say it is worse than the zero-value punishment for murder, which is some considerably more severe punishment.

¹⁰ This is not the only possible justification for this view. Even if the desert-goodness of inflicting a punishment on someone who ideally deserves it were equal to the desert-evil of inflicting it on an innocent person, a retributive theory could distinguish between doing and allowing and say the state's causing the evil is more objectionable than its failing to cause the good.

¹¹ D. Dolinko, “Justice in the Age of Sentencing Guidelines,” *Ethics* 110 (2000): 563–85.

would not, this will be an improvement not only in individualistic but also in holistic terms. But if we think seriously about individualistic retributivism, we cannot believe that there is an exactly optimal punishment for a given crime—say, an exact number of years, months, and days in prison. Even given all the details of the crime, there is at best a vague band of punishments none of which is less ideal than the others, and this band may be fairly wide, so that for a given crime it runs from, say, one year in prison to two. But then judges who have sentencing discretion may without departing from individualistic standards impose quite different punishments for the very same crime, some choosing a sentence near the top of the one- to two-year range while others choose one near the bottom. This will be a holistic injustice, since it treats similar cases differently. It is here that the holistic principle gives its distinctive support to uniform sentencing. Because individualistic retributivism is necessarily vague, even judges applying it faultlessly can introduce discrepancies in sentencing that are by holistic standards unjust.

The second implication of the holistic principle is to support certain arguments for leveling down. The most prominent of these have been used to oppose the death penalty in the one Western country that still uses it, the United States. One argument claims that the application of the death penalty is racially discriminatory, so that, other aspects of their crimes held constant, black murderers are more likely to receive the death penalty than white murderers. A second argument, which is a generalized version of the first, holds that the application of the death penalty is arbitrary, with no discernible standards determining which murderers are put to death and which are not. Together these arguments led the U.S. Supreme Court in 1972 to declare that the country's existing death penalty statutes, which gave juries complete discretion in deciding whether to impose the death penalty, were unconstitutional. Four years later the same court held that statutes revised to identify specific aggravating factors that juries must find in order to impose the death penalty were constitutional. But many American opponents of the death penalty believe the arguments still have force. They believe the application of the death penalty remains discriminatory and capricious and for that reason should be abolished.

These arguments would not favor leveling down if they held only that discriminatory or arbitrary procedures are imposing the death penalty on offenders who do not deserve it, for example, on black murderers whose crimes do not have the relevant aggravating features. Though doubtless important to many death-penalty opponents, this consideration is hard to introduce in legal argument, since it requires showing that individual juries have applied the death penalty wrongly. As a result, the arguments tend to be stated in a way that implies only that some who do deserve the death penalty are escaping it. Consider the following from Justice William O. Douglas's opinion in the 1972 Supreme Court decision: "A law that stated that anyone making more than \$50,000 would be exempt from the death penalty would plainly fall, as would a law that in terms said that blacks...or those

who were unpopular or unstable should be the only people executed.”¹² By describing only discriminatory *exemptions* from punishment, these analogies imply versions of the anti-death-penalty arguments that favor leveling down and therefore require a holistic principle. Of course, just as some philosophers reject desert leveling down in general, so some hold that retributive desert is governed only by individualistic principles. Ernest van den Haag has taken this line, saying that “the guilty do not become innocent or less deserving of punishment because others escaped it.”¹³ But I again think the most intuitive version of retributivism includes not only an individualistic principle but also one concerned with the pattern of punishments as a pattern, and that principle can favor leveling down.

It may be objected that our commitment to the holistic principle is not consistent. Though we appear to favor some cases of leveling down, we surely would not favor leveling up, or giving some people more punishment than they individually deserve because others are receiving more. Thus, if a society had imposed an intrinsically excessive punishment such as the death penalty for rape and was applying it discriminatorily, so that only black rapists were being executed, we would not favor executing white rapists as well. There are two replies to this objection. First, even if leveling up is always on balance wrong, the discriminatory application of an excessive punishment involves an evil that is not present where the same punishment is applied evenhandedly. Second, the details of individualistic retributivism can explain why leveling up is always on balance wrong. In Figure 9.1 the desert curves slope down more steeply outside their peaks than inside them—that is, more steeply when a person is getting more than he ideally deserves than when he is getting less.¹⁴ This implies, plausibly, that it is worse in desert terms if a person receives a fixed amount more punishment than is optimal than if he receives the same amount less. This in turn creates an asymmetry between leveling up and leveling down. In both cases the holistic gain from leveling may be the same, but in leveling up the individualistic loss is greater. In fact, if the slope of the curves outside their peaks is sufficiently steep, the individualistic loss from leveling up will always outweigh the holistic gain, so leveling up is always on balance wrong whereas leveling down is sometimes right.

A different objection, due to Dolinko, says the holistic arguments are too strong, because if successful they rule out not only the death penalty but any punishment

¹² *Furman v. Georgia*, 408 U.S. 239 (1972), quoted in H. A. Bedau (ed.), *The Death Penalty in America*, 3rd ed. (New York: Oxford University Press, 1982), pp. 255–56. Note that Douglas’s remarks do not concern only the application of existing law but also the formulation of the law. In “Justice and Desert in Liberal Theory” Scheffler allows that a purely individualistic retributivism can hold that the discriminatory application of existing penal norms is unjust (pp. 986–87n.). But the anti-death-penalty arguments go beyond that concern to oppose disproportions in the pattern of punishments no matter how they arise.

¹³ E. Van Den Haag, “In Defense of the Death Penalty: A Legal-Practical-Moral Analysis,” *Criminal Law Bulletin* 14 (1978): 56.

¹⁴ This is not an essential feature of the graphs, and in particular is not required by proportionality. I have included it because of its attractive implications for retributivism.

whatever. Imagine that, because the application of the death penalty is discriminatory, we replace it with a twenty-five-year prison term. The application of this punishment, too, will presumably be discriminatory, with black defendants being assessed the full twenty-five years while whites receive, say, only twenty. The same holistic argument now requires us to replace the twenty-five years with twenty, and, assuming the discrimination continues, to keep replacing more with less severe punishments until no punishment remains. But surely an argument that forbids any punishment whatever is too strong to be acceptable.¹⁵

This objection is again answered by the details of the individualistic principle, this time by the fact that the curves in Figure 9.1 are curves, whose slope increases as they run down from their peaks toward the origin. This implies that the first unit of severity by which a punishment falls short of the optimum involves a comparatively small loss of individualistic desert-value, the second a larger loss, and every subsequent unit a larger loss still. This in turn implies that the holistic gain from leveling down may outweigh the individualistic loss when the resulting punishment is a little less severe than the optimum but will be outweighed when that punishment is much less severe, thereby stopping Dolinko's regress in a principled way.

A final objection says there are other disproportions that we do not take to be unacceptable or to warrant the abolition of severe punishments. So long as society does not spend all its resources on policing, there will be some murders that are not solved, so some murderers receive no punishment, and others in which the specific evidence that would justify the death penalty is not found. Yet no one thinks the resulting disproportions, though large, justify abolishing the death penalty.

There are several possible replies to this objection. First, in this case eliminating the disproportion would have a significant cost, namely whatever other good could be done with the funds transferred to policing, whereas eliminating discrimination or arbitrariness has no such cost. Second, our thinking about the criminal justice system may object more to evils the state causes than to ones it merely allows, and a disproportion that results from failing to spend more on policing is one the state allows, whereas one that results from discriminatory or arbitrary procedures is one it causes. Finally, we may care about proportionality not only in the final pattern of punishments but also in people's prior likelihoods of suffering punishment. If the shortfall in policing is spread evenly among sectors of society, rather than being concentrated, say, among blacks, these prior likelihoods will not be disproportionate: everyone will face the same probability of suffering the death penalty if he murders. This distinguishes the shortfall case from that of discrimination, in which black murderers have a higher prior likelihood of suffering the death

¹⁵ D. Dolinko, "How to Criticize the Death Penalty," *Journal of Criminal Law and Criminology* 77 (1986): 575-77, 582-83.

penalty. It does not, however, distinguish the shortfall case from that of simple arbitrariness, since there, too, prior likelihoods are the same. This may be one reason why the more abstract arbitrariness argument against the death penalty seems less powerful intuitively than the discrimination argument. In the one case the disproportion is confined to the final pattern of punishments; in the other it also affects people's prospects of suffering punishment.

I will assume that some combination of these and perhaps other arguments answers this last objection to retributive leveling down. If so, then despite differing from moral desert in several aspects of its content and structure, retributive desert combines individualistic and holistic elements in similar ways and to similar effect. It, too, therefore, is not purely individualistic but contains an important holistic component. The same is even more true, I will now argue, of economic desert.

III. Economic Desert

By economic desert I mean desert of economic goods, and, more specifically, of income for a given time's work, say, an hour's. There are many theories of economic desert, but the most prominent make a person's desert depend on either or both of two factors: her contribution and her effort. By contribution I mean the degree to which her work contributes to satisfying the preferences of others, where this is understood in a marginalist and counterfactual way. It is her contribution given the work of others, as well as given the preferences of others, the resources available to her, and so on. And it is measured by comparing how far people's preferences are satisfied when she works with how far they would be satisfied if she did not work and other workers made appropriate adjustments: that difference is her economic contribution. Effort I understand, somewhat expansively, as everything negative about work or contrary to the worker's preferences: this includes the work's strenuousness, its unpleasantness, the amount of training it requires, and any health risks or other hardships it imposes. Some theories base a person's economic desert only on her contribution or only on her effort, but in my view the most attractive theory makes it depend simultaneously on both. According to this composite theory, the income a person deserves for an hour's work is determined jointly by the benefits it provides to others and the costs it involves for her, and the theory is especially attractive if it combines these elements in a multiplicative way. Having generated separate measures of a person's contribution and effort, it determines her overall desert not by adding but by multiplying them. The main merit of this approach is that it yields the right result when a person makes either no contribution or no effort. If she exerts effort in an activity that contributes nothing to others, such as counting blades of grass, an additive theory, like one that values only effort, must say she deserves some income for her work. But a multiplicative theory says that if her contribution is zero, her overall desert is zero. Similarly, if she benefits others in a way that involves no cost to herself, say, because her skin

painlessly emits a pleasant scent, a multiplicative theory again says her overall desert is zero. Because these claims are attractive, so is the multiplicative theory that yields them.

Whichever it values of contribution and effort, an economic theory differs systematically in its content from a theory of moral desert. On the one side, the plausible bases of economic desert are independent of virtue. If a person contributes to others or does unpleasant work, it does not matter if his motive is an altruistic desire to benefit others or greed for a higher income; so long as he does those things, he deserves income for them. On the other side, what he deserves on the basis of contribution or effort is only income and not happiness; if he does not derive joy from his earnings, that is no concern of economic desert. These differences illustrate an important point of Feinberg's: that what is deserved and its desert-base must be appropriate to each other.¹⁶ In this case the moral and economic theories each assign value to combinations of states with the same type of value. The moral theory says the intrinsic value of pleasure or pain is deserved on the basis of the intrinsic value of virtue or vice; the economic theory says the instrumental value of income is deserved on the basis of the instrumental good of contributing to others or the instrumental evil of effortful work. In each case what is deserved fits its desert-base by being a value of the same type: intrinsic for intrinsic or instrumental for instrumental. In his main discussion of economic desert Rawls directs the bulk of his attention to the view that people deserve income on the basis of their moral virtue.¹⁷ As many critics have noted, this is not a view defenders of economic desert actually hold. There is a reason: the view Rawls discusses ignores Feinberg's point by mixing different types of value. What people deserve on the basis of virtue is not money but happiness; what makes them deserve money is not virtue but the instrumental qualities of contribution and effort.

An economic theory also differs structurally from one about moral desert, though in the contrary way to a retributive theory. Whereas retributivism concerns only negative desert, making no claims about deserved rewards, the economic theory concerns only positive desert, or what people deserve for positive contributions or efforts. If someone is economically destructive, she may be liable to criminal punishment or owe damages but does not deserve a negative income; similarly, if her work is the opposite of effortful, that is, enjoyable, she deserves on that basis only no and not a negative wage. So graphs for economic desert again mirror only parts of the moral graphs, though different ones from those for retributive desert. Thus, the theory's holistic principles define their ideal distributions only in the top right quadrant of Figure 9.2, which now represents distributions between people who have both made positive contributions or positive efforts. But

¹⁶ Feinberg, "Justice and Personal Desert," in his *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, N.J.: Princeton University Press, 1970), p. 73.

¹⁷ J. Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), pp. 310–15.

the theory again retains formal features of the moral graphs, in particular their valuing of proportionality, so its holistic principles assign their highest value to distributions in which the ratio of people's incomes equals the ratio of their contributions or efforts. And if the theory is multiplicative, it also values proportionality in people's overall economic desert. Imagine that someone who previously deserved \$10 an hour for contribution and \$10 an hour for effort doubles her contribution per hour while leaving her effort unchanged. An additive theory says her overall desert has increased from \$20 an hour to \$30, which is not a doubling. But a multiplicative theory says that, other things equal, any doubling of contribution or effort results in a doubling of the total income deserved.

Especially in this multiplicative form, the composite theory allows a partial defense of the incomes generated in a free market. These incomes result from the joint operation of demand and supply in the labor market, which respond roughly to people's contributions and efforts. On the one side, the more an employer thinks an employee will promote his, the employer's, ends, the more he will pay for the employee's labor; on the other side, the more strenuous the labor is or the more training it requires, the more an employee will insist on being paid to do it. Contribution and effort are therefore rewarded on, respectively, the demand and supply sides of the labor market, and they are also rewarded in a multiplicative way, since the market pays nothing for contribution-less efforts or effortless contributions. This desert-based defense of the market is only partial, however. A complete theory of distributive justice may contain principles other than desert ones—for example, principles about need—and they may condemn distributions the desert theory on its own approves. Even that theory itself does not approve every outcome of the market. As Robert Nozick points out, the market allows unilateral transfers such as gifts and bequests that do not reward contribution or effort.¹⁸ In addition, the market rewards perceived rather than real contribution and effort, and gives greater rewards for contributions to or efforts from the wealthy even when those are not in themselves greater. In these and other ways market outcomes can depart from what economic desert approves, but, following Nozick, we can say that distribution in accordance with contribution and effort is an important 'strand' in market distribution, which is justified by desert principles to the extent that it contains it.

In so justifying the market, the economic theory allows that a person's income may legitimately be affected by factors that are beyond his control and therefore matters of luck. This is clearly true of the contribution principle, since how much a person benefits others can be influenced by his natural talents, how those talents fit others' preferences, and how common they are in his society. It is also true, though on more subtle grounds, of the effort principle.¹⁹ The theory therefore

¹⁸ R. Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 157.

¹⁹ I here agree with the conclusion of George Sher's "Effort and Imagination," though using a different understanding of effort and relying on different arguments.

violates a condition some philosophers apply to all distributive principles, namely that income differences are justified only if they reflect differences in people's choices or in features all had a fair opportunity to acquire. I will not try to decide the adequacy of this responsibility condition here. Though it is accepted by many writers, it is rejected by others and, as empirical studies have shown, is also rejected in everyday thinking about distributive justice.²⁰ But I will suggest two reasons why the economic theory is not simply and obviously objectionable on this score.

First, the responsibility condition cannot be used to justify Rawls's distinction between economic and retributive desert, since retributive desert, too, is affected by luck. Whether a person deserves any legal punishment can depend on whether he happened to be restrained from acting on a murderous intention, and how much punishment he deserves can depend on whether his bullet hit its target. But if desert of outcomes as serious as execution and imprisonment can depend on luck, why not also desert of incomes?²¹ Second, the economic theory does not simply ignore responsibility. It does or can say that people deserve income only on the basis of contributions or efforts they chose to make, so choice is a necessary condition for these desert bases though it does not determine their content entirely. The theory is therefore utterly different from ones that say people deserve just on the basis of having talents or belonging to a certain social caste. It may be objected that if choice is only part of what determines a person's contribution, he should deserve only on the basis of that part and not for his contribution as a whole. But here the theory can answer using marginalist concepts: if a person has talents, his choosing to exercise them makes all the difference between their contributing to others and their not doing so, so he is appropriately rewarded for his whole contribution. A similar point applies to effort. A person's effort can be affected by luck if he has the option of doing more or more strenuous work than others can—for example, if his factory offers overtime where other workers' factories do not; this is an advantage to him if justice allows the net effect of his doing that work and being paid for it to make him better off. But, again, if his choosing to do the work makes all the difference between his expending that effort and not, the theory can say he is appropriately rewarded for all his effort.

²⁰ For a summary see D. Miller, *Principles of Social Justice* (Cambridge, Mass.: Harvard University Press, 1999), ch. 4.

²¹ Some who defend the responsibility condition distinguish between option luck, or luck in the outcomes of gambles people choose, and brute luck, or luck in the circumstances in which they choose; only differences resulting from brute luck, they say, are unjust (see, e.g., R. Dworkin, "What is Equality? Part 2: Equality of Resources," *Philosophy and Public Affairs* 10 [1981]: 283–345). They can accept differences in retributive desert between those whose bullets do and do not hit their targets, saying these are a matter of option luck. But it is harder to see how they can accept the more fundamental difference between those who are and are not restrained from acting on a murderous intention. Especially if this is a difference between those who do and do not have bystanders present, is it not a matter of brute luck? At the very least, the question of whether retributive desert can be affected by luck is sufficiently complex that there is no simple way of justifying Rawls's distinction by appeal to responsibility.

With these preliminaries behind us, let us turn to the role in economic desert of individualism and holism. Here my claim is that the most plausible economic theory contains only holistic and no individualistic principles. It cares that the distribution of incomes across individuals be proportioned to their contributions and efforts but says nothing about what those individuals on their own deserve. This claim is not original. It is implicit, I think, in Nozick's description of principles mandating distribution in accordance with virtue or economic contribution as "patterned principles," which care about the pattern of distribution in society as a whole, and especially clear in his assumption that these principles are indifferent between instantiations of the same pattern at different levels.²² The claim is also made at several points by Feinberg. In *Social Philosophy* he argues that distributive justice is an instance of comparative justice, which is his name for holistic justice, and includes within it principles rewarding contribution and effort.²³ And at least one passage in his essay "Noncomparative Justice" takes a similar line, rejecting the suggestion that a worker's low wage may be less than she deserved considered on her own by saying that all claims about fair wages turn on comparisons between her income and those of other workers or of her employers.²⁴ This claim that economic desert is only holistic is the one I want to defend.

The principal ground for this claim is simply intuitive: that individualistic claims about economic desert are not plausible. Imagine that a person makes a fixed contribution to others, in either absolute or percentage terms, or exerts a fixed amount of effort. How much income he receives for doing so will depend on the level of development of his economy. If his economy is not very productive, he will receive only a low income; if it is technologically advanced, he will receive much more. An individualistic principle of economic desert has to say that one of these outcomes has more desert-value, but I do not see that this is so. In each economy there are facts about what other people earn for similar contributions or efforts, and if his income is out of line with theirs this is unjust. But this claim follows from holistic principles, and I do not see that any of the additional claims individualistic principles support are true. If both the low incomes in the unproductive economy and the high incomes in the productive one are divided proportionally, I see no difference in desert-value between them. Of course, many will say the productive economy is preferable, but here other values may be playing a role. In particular, people in the productive economy may be happier, which is both

²² Nozick, *Anarchy, State, and Utopia*, pp. 155–60, 209–10.

²³ Feinberg, *Social Philosophy* (Englewood Cliffs, N.J.: Prentice-Hall, 1973), pp. 98–9.

²⁴ Feinberg, "Noncomparative Justice," pp. 278–9. Nor are purely holistic claims about desert original to these 20th-century philosophers. Consider Aristotle's formula for distributive justice, which requires the ratio of A's reward to B's to equal the ratio of A's merit to B's (*Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes [Princeton, N.J.: Princeton University Press, 1984], 1131a39). This formula, which Aristotle applies primarily to the distribution of political offices, can be equally satisfied at different levels and is therefore also purely holistic.

good in itself and may better reward their virtue. But we can eliminate these considerations by imagining that people in the two economies are exactly equally happy. Now there is surely no value difference between them, which implies that the only economic principles relevant to them are holistic. As Feinberg notes, the situation here is entirely different from that of retributive desert. Of two systems assigning punishments in a perfectly proportional way, one can be worse in retributive terms because it assigns punishments that are intrinsically too severe, such as capital punishment for car theft, or too lenient.²⁵ But the parallel claim is not plausible for economic desert, where proportional distributions at very different levels of income seem equally just.

The holism of economic desert is illustrated by several aspects of contemporary practice and attitudes. Consider how incomes change in professional sports. Starting from a stable salary structure, one player, usually a top star, receives a salary increase, often when one team lures him away from another. Other stars then demand comparable increases, saying that since they contribute as much to their teams they deserve to be paid as much, and proportional increases follow for lesser players. What drives this process is not a concern with absolute levels of income; though everyone would like to be paid more, no one makes a desert-based complaint about his place in the initial structure. Instead, the driving force is a concern with comparative incomes, and especially that equally proficient players be paid equally. Nor is this type of concern limited to high earners such as athletes. In the late 1970s the Labour government in the United Kingdom initiated a “social contract” with the country’s labor unions whereby both would restrain wage increases in order to combat inflation. But this contract was opposed by some union members on the ground that the resulting restraints were eroding “wage differentials.” Their concern was not with the absolute level of their income but with whether proportional differences in contribution and effort were being proportionally rewarded.²⁶

Or consider the practice of assessing jobs for their “comparable worth,” which can result in some workers’ incomes being adjusted upwards if they are found to be underpaid compared to others. This practice, which is often said to illustrate the relevance of desert to economic justice,²⁷ usually assumes a composite theory valuing both factors affecting contribution and ones affecting effort. But the impor-

²⁵ Feinberg, “Noncomparative Justice,” p. 279.

²⁶ An interest in comparative incomes can reflect not only a concern with holistic desert but also a concern with status, since earning more than others confers high status and earning less confers low status (see R. H. Frank, *Choosing the Right Pond: Human Behavior and the Quest for Status* [New York: Oxford University Press, 1985]). But the two concerns come apart in the following case. Starting from a stable salary structure, the top contributor increases his contribution while the contributions of others remain unchanged. Workers concerned only with status would view a salary increase for the top contributor unaccompanied by any increase for themselves only negatively, since it would only increase their status inferiority, whereas ones concerned with desert would accept it. I think many workers, e.g., many professional athletes, would accept an increase in this case.

²⁷ See, e.g., Miller, *Principles of Social Justice*, p. 84; and O. McLeod, “Desert and Wages,” in Pojman and McLeod, *What Do We Deserve?* p. 271.

tant point is that it assesses only jobs' *comparable* worth. It does not attempt to determine the ideal wage for a given job considered on its own, but looks only at how that job's wage compares to those for other jobs, given their contribution and effort. It asks whether certain workers are underpaid not in absolute terms but only in comparison with others. Nor in my view is the reason for this only epistemic. It is not that comparable-worth assessors would determine ideal wages for jobs considered on their own if they could; instead, their practice implicitly recognizes that, in the economic realm, the only relevant desert principles are holistic.

Empirical studies of popular opinions about distribution point in a similar direction. They find strong support for ideas about economic desert, in fact stronger than for any competing ideas on this topic. But the questions that elicit this support are consistently phrased in comparative terms, asking, for example, whether "people who work hard deserve to earn more than those who do not."²⁸ A positive answer to this question does not unequivocally indicate a concern for patterns as patterns; it can also follow from an individualistic principle that satisfies the proportionality condition. But the fact that researchers ask only comparative questions and none about what individuals deserve on their own suggests that they recognize, even if only implicitly, that what will win support is only a holistic claim about the relations between people's distributive shares.

Though this holistic view has been defended by Nozick and Feinberg, other prominent accounts of economic desert are individualistic. One view holds that people deserve income on the basis of their effort as compensation for the hardship of exerting that effort; another holds that what people deserve on the basis of their contribution is an income equal to their marginal economic product. These views are individualistic because they consider each person's effort or marginal product on its own, without comparing it to other people's. They also have an optimality structure, since they hold that for each person there is a specific income that is ideally deserved, either one that exactly compensates him for his effort or one that exactly equals his marginal product. In a longer chapter I would present several objections to each of these views; here I will describe only the main difficulty facing each.²⁹

²⁸ A. Swift, G. Marshall, and C. Burgoyne, "Which Road to Social Justice?" *Sociology Review* 1 (1992): 29; and G. Marshall *et al.*, "What Is and What Ought to Be: Popular Beliefs about Distributive Justice in Thirteen Countries," *European Sociological Review* 15 (1999): 356–57, 366.

²⁹ The two views tend to be held on their own rather than as part of a composite theory like the one I have described. I think there is a reason for this. If the individualistic principles hold that a person ideally deserves \$10/hour for his contribution and \$10/hour for his effort, an additive theory can say he deserves \$20/hour overall. But if he receives \$20/hour, how can we know that he is receiving \$10 for contribution and \$10 for effort rather than \$15 for contribution and \$5 for effort or even \$20 for contribution and \$0 for effort, each of which is less good in desert terms? (There are comparable difficulties in a multiplicative theory.) I am not saying there is a general difficulty about combining measures of contribution and effort into a measure of overall economic desert with an optimum. But there are difficulties if the initial measures define optima for contribution and effort on their own. Because the compensation and marginal-product views do this, I am suggesting, it is no surprise that they are usually held on their own.

The compensation view was proposed by Feinberg in his 1963 paper “Justice and Personal Desert,” though it seems absent from his later writings; versions of it are also defended by Wojciech Sadurski, George Sher, and Julian Lamont.³⁰ The main objection to it is that it makes justice require that the net effect on a worker’s welfare of expending effort and being paid for it leave her no better off than if she had not worked. This means, first, that the view does not allow positive incentives for effortful work; though it removes negative incentives, it forbids any positive inducements to unpleasant labor. This makes it hard to see how an economy governed by the view could ever develop technologically. Why would workers exert the effort needed to make their economy more productive if doing so never gave them a welfare boost? Second, this feature of the compensation view is utterly at variance with common sense. Common sense would find the proposition that justice requires the net effect of working and being paid to leave one no better off than if one had not worked absurd, and it would likewise find absurd a more restricted application of this view to jobs involving above-average effort. Most people agree that if a medical career requires extra years of training, it is appropriate that it be paid more highly. But they show no concern that the extra payment do no more than exactly compensate for the cost of the training; they allow that the net effect of training and being paid for a medical career can make one better off. This is because, as I have argued, the intuitive effort principle is only holistic, caring only that those who expend more effort earn more and saying nothing about what individual efforts on their own deserve.

The marginal-product view is defended by David Miller³¹ and has at least one notable advantage. It avoids my main argument against an individualistic contribution principle by changing the measure of contribution from how much a person satisfies others’ preferences to how much she produces purely economic goods. This allows the view to yield consistent results in different economies. In an unproductive economy a person’s marginal product is small, as is the income she deserves for it, but in an advanced economy both her product and the income that matches it are larger. When contribution is measured economically, the same individualistic principle yields appropriately different results in differently developed economies.

But this advantage comes at the cost of preventing the view from giving a non-circular justification of the incomes generated by the market. Imagine, to take the clearest case, that one person provides a service to another, say, by playing music for her or giving her a massage. I do not see how the marginal-product view can

³⁰ Feinberg, “Justice and Personal Desert,” pp. 92–94; W. Sadurski, *Giving Desert its Due* (Dordrecht: Reidel, 1985); G. Sher, *Desert* (Princeton, N.J.: Princeton University Press, 1987), pp. 99–108; and J. Lamont, “Incentive Income, Deserved Income, and Economic Rents,” *Journal of Political Philosophy* 5 (1997): 26–46.

³¹ D. Miller, *Market, State and Community: Theoretical Foundations of Market Socialism* (Oxford: Clarendon Press, 1989), ch. 6.

assign an economic value to this service other than by equating it with what the person is actually paid for it. But then the claim that the market is just because it pays people what they deserve says nothing more than that the market pays what the market pays: if there is no independent measure of contribution, there can be no independent justification of market distribution. The same problem arises in a more complex case where several people cooperate in an enterprise that generates a certain total income. Here the marginal-product view seems to be making an independent claim when it says each worker should be paid an amount equal to the difference her labor makes to that total. But if we ask why the proper sum to be divided among the workers is the total income they generate, the answer again is that the market measures their joint contribution, which again means the market is being justified for paying what it pays. No comparable circularity arises if contribution is measured as I have suggested, by how much a person satisfies others' preferences. That measure is conceptually independent of the market and can underwrite an independent, if only partial, justification of it. But given that measure, the only plausible contribution principle is holistic.

When the distinction between individualistic and holistic principles is attended to,³² I believe it becomes evident that the most plausible principles of economic desert care only that incomes be proportioned to people's contributions and/or efforts and not about the absolute level at which this is done. This is a first respect in which economic desert is more holistic than moral or retributive desert, but this is not because it adds a holistic element not present in those theories. On the contrary, economic desert differs only because, while sharing a holistic element with those theories, it omits an individualistic element they contain. And there is another respect in which economic desert is more holistic.

Any holistic desert principle values a situation in which the pattern of distribution of one state such as punishment or income matches the pattern of distribution of another that is its desert-base. But in many holistic principles this desert-base is individualistic, involving states of individual people apart from any relations to similar states of others. This is certainly true of moral desert, where the desert-base of virtue is on many views a purely internal state of a person. A holistic moral principle cares about the distribution of virtue across people, but what contributes to this distribution is in each case a state of an individual considered on his own.

³² This is not always done, even by writers who draw the distinction, such as Feinberg. Though his principal view in *Social Philosophy* is that the principles of economic desert are only holistic, he defends a contribution principle in part by analogy with the idea that items a person owns and has loaned out should be returned. He writes, "the return of contribution is not merely a matter of merit deserving reward. It is a matter of a maker demanding that which he has created and is thus properly his" (p. 116). But Feinberg immediately restates the principle as requiring that the ratio of X's share of income to Y's equal the ratio of X's contribution to Y's, and these two statements are entirely different. The idea of returning what a maker has made is individualistic, implying that for each person a particular income is deserved independently of what other people deserve. But a ratio principle is holistic and can be equally well satisfied at different levels of income.

The same is true of retributive desert. The seriousness of a person's crime is not a fact just about him, since it can depend on how much harm he caused his victim. But to assess the seriousness of one crime we need not consider facts about other crimes; the desert-base here is independent of relations to other states of the same type. The same holds for effort in economic desert, since how many negative effects his work has on a person is a fact just about him. But it is not true of contribution if that is understood in a marginalist way. If a person's contribution is the difference his work makes to the satisfaction of others' preferences given the work of others, then the assessment of his achievement of this desert-base refers implicitly to the achievements of everyone else. Even if his work is held constant, the absolute contribution it makes, which is what gets compared with others' contributions in the initial holistic pattern, can change as their work changes. It is not a fact just about him or about him and the few people he directly affects; his contribution is defined as the difference his labor makes to overall preference-satisfaction given the labor of everyone else. There is therefore a second respect in which economic desert is more holistic than moral and retributive desert: in one (though not the other) of its holistic principles, the achievement by any one person of the relevant desert-base necessarily involves relations to similar facts about others. This principle is therefore holistic not only in its form, for valuing a pattern of distribution across individuals, but also in its content, since the state that enters into its initial pattern refers implicitly to similar states of all people. Since the holistic principles in moral and retributive desert do not have this feature, economic desert is in a second respect more holistic than they.

IV. Scheffler On Holism and Desert

Let me conclude by relating these points to Scheffler's argument about holism and economic desert. Scheffler identifies several respects in which distributive justice is holistic. If my arguments are sound, the most plausible theory of economic desert is also holistic in those respects. The theory is not inappropriate to but precisely fits the holism of the distributive realm.

Some of Scheffler's claims concern what I have called the form of distributive principles, or the type of state they value. Thus, he says "the justice of any assignment of economic benefits to a particular individual always depends—directly or indirectly—on the justice of the larger distribution of benefits in society," and quotes approvingly Rawls's remark that distributive principles do not apply to "a single transaction viewed in isolation."³³ But a theory of economic desert that values only patterns agrees, saying there is no desert-value in a situation involving

³³ Scheffler, "Justice and Desert in Liberal Theory," p. 984; the Rawls quote is from *A Theory of Justice*, p. 87.

only one person. Even a retributive theory agrees to some extent, since it says we cannot know conclusively whether a given punishment for a person is just unless we know how other people are being punished. But the economic theory goes further, denying that any desert claims whatever can be made about a single person; it is not only partly but wholly holistic. The particular form this theory's holism takes appears to differ from that of Rawls's. Whereas the economic theory evaluates patterns of distribution, Rawls's theory evaluates the institutions comprising the "basic structure" of a society and says that if these are just, any distribution they produce is also just; this point is one Scheffler sometimes emphasizes. But I am not persuaded that there is a significant difference here, since Rawls's judgments about institutions turn ultimately on the distributions they produce. And even if there is, it is not a difference between one theory that is formally holistic and another that is not; at best, it is a difference between theories that express their holism in slightly different ways.

Others of Scheffler's claims concern what I have called the content of distributive principles, or whether the states of people they take to ground distributive claims involve relations to other people. Thus, he says people's economic contributions are "interconnected" because "each person's capacity to contribute depends on the contributions of others"; the economic value of their talents is socially determined because "it depends both on the number of people with similar talents and on the needs, preferences, and choices of others"; and their prospects are linked because "any decision to assign economic benefits to one person or class has economic implications for other persons and classes." Given these interconnections, he concludes, principles assigning benefits to people on the basis of individualistic facts about them make "no normative sense."³⁴ But these claims are not relevant to an effort-based desert principle, since, as I have argued, how much effort a person expends is a fact just about him. And a contribution principle entirely accepts the claims. Especially if it defines contribution in a marginalist way, it agrees that each person's contribution is affected by the contributions of others, as well as by the talents and tastes of others; it characterizes contribution precisely so as to give those connections weight. A defender of the principle may not see the interconnections as grounding a positive argument for it—nor suggest, as Scheffler sometimes does, that they are more important in complex modern economies than in primitive ones. Instead, she may simply propose contribution to others as in itself the intuitively correct basis for economic desert and argue that its best characterization has always been marginalist. Even so, the desert theory she defends does not ignore the facts about interconnectedness Scheffler cites; instead, it represents one way of giving them normative weight.

Scheffler's arguments raise important issues. Many contemporary philosophers reject economic desert on broadly Rawlsian grounds, but either do not reject

³⁴ Scheffler, "Justice and Desert in Liberal Theory," p. 985.

retributive desert or think different arguments are needed to refute that view. If this position is sound, however, there must be an argument that rules out economic desert but does not generalize to retributive desert. Scheffler proposes an argument contrasting the allegedly individualistic character of desert with the holism of distributive justice. In response, I have argued that even moral and retributive desert contain holistic elements and that economic desert is thoroughly holistic. If these claims are correct, economic desert precisely fits the holism of the distributive realm. The argument separating distributive from retributive desert has yet to be found.

Virtuous Act, Virtuous Disposition

Everyday moral thought uses the concepts of virtue and vice at two different levels. At what I will call a global level it applies these concepts to persons or to stable character traits or dispositions. Thus we may say that a person is brave or has a standing trait of generosity or malice. But we also apply these concepts more locally, to specific acts or mental states such as occurrent desires or feelings. Thus we may say that a particular act was brave or that a desire or pleasure felt at a particular moment was malicious. Even when they concern acts, these last judgments are of virtuousness rather than of moral rightness. They therefore turn essentially on a person's motives; while he can act rightly from a bad motive, he cannot act virtuously from a bad motive. But they assess the virtue or vice of particular acts and mental states rather than of persons or traits of character.

These global and local uses of the virtue concepts are clearly connected, in that we expect virtuous persons to perform and have, and virtuous traits to issue in, particular virtuous acts, desires, and feelings. A philosophical account of virtue should explain this connection, but there are two different ways of doing so. Each takes one of the two uses to be primary and treats the other as derivative, but they disagree about which is the primary use.

A *dispositional* view takes the global use to be primary and identifies virtuous acts, desires, and feelings in part as ones that issue from virtuous dispositions. Aristotle famously took this view. In the *Nicomachean Ethics* he said that for an act to be virtuous it must meet some initial conditions, including about its occurrent motivation, but must also “proceed from a firm and unchangeable character”; if it does not, it may be such as a brave or generous person would perform, but is not itself brave or generous.¹ The dispositional view requires that we be able to identify

¹ *Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984), 1105a31–b12.

virtuous character traits and persons independently of virtuous acts or feelings; otherwise its account of the latter will be circular. But assuming some such identification, it treats virtuous dispositions as primary and defines virtuous occurrent states derivatively, as ones that proceed from such dispositions.

A contrary *occurrent-state* view takes the local use to be primary and identifies virtuous dispositions as ones to perform virtuous acts and to have virtuous desires and feelings. W. D. Ross took this view. In *The Right and the Good* he held that virtuous action is action from any of three motives: the desire to do one's duty, the desire to bring into being something good, and the desire to produce some pleasure, or prevent some pain, for another being.² Since these desires are occurrent states, Ross applied the virtue concepts first to such states and then defined virtuous dispositions derivatively, so bravery is a disposition to brave acts, desires, and feelings, and generosity a disposition to generous ones. This view must be able to identify virtuous occurrent states independently of virtuous dispositions, but Ross's discussion shows one way to do so. The virtuousness of an act depends on the motive or desire it is done from, rather than on its effects or conformity to principles of duty. And the virtuousness of a desire or feeling depends on its appropriateness to the moral value of its object. Thus, desiring or taking pleasure in something good for another person, such as his pleasure, for its own sake is virtuous and in particular generous, while desiring or taking pleasure in an evil such as his pain is malicious.

The dispositional and occurrent-state views agree that virtuous dispositions tend to issue in virtuous acts, desires, and feelings, but they give different explanations of this fact. In consequence they also disagree about some particular cases. Imagine that a person performs an act from what Ross would call a virtuous motive but does not have a stable disposition to act from that motive; imagine, for example, that he promotes another's pleasure from an occurrent desire for that pleasure for its own sake but does not normally have such desires and therefore now acts out of character. The occurrent-state view implies that his act is virtuous and in particular generous; the dispositional view implies that it is not. For this reason, the two views also disagree about how tightly the global and local uses are connected. Both hold that virtuous dispositions tend to issue in virtuous occurrent states, but while the dispositional view also holds that virtuous states necessarily issue from virtuous dispositions, the occurrent-state view rejects that converse claim.

In the contemporary virtue ethics literature the dispositional view seems overwhelmingly dominant. Many virtue ethicists give formal definitions of virtue that concern only character traits and not particular acts and feelings; if they discuss

² *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 161–62. Judith Jarvis Thomson also endorses the occurrent-state view in Harman and Thomson, *Moral Relativism and Moral Objectivity* (Oxford: Blackwell, 1996), pp. 144–47 and Thomson, “The Right and the Good,” *Journal of Philosophy* 94 (1997): 280; I defend it briefly in *Virtue, Vice, and Value* (New York: Oxford University Press, 2001), pp. 42–44.

the latter, it is at best secondarily.³ Others emphasize how their view focuses on character rather than on discrete acts considered apart from character.⁴ Most explicitly, Julia Annas says the virtues are essentially dispositions to act and therefore connected to judgments about the value of one's life as a whole; on that basis she rejects views like Ross's, which locate virtue initially in occurrent attitudes, as giving an "idiosyncratic" and "reduced" picture of virtue.⁵ But it seems to me that, whatever philosophers may say, the contemporary commonsense understanding of virtue is clearly the occurrent-state one. When everyday moral thought applies the virtue concepts, it is primarily to occurrent states considered on their own.

Imagine that, walking down the street, you see someone kick a dog from an evident desire to hurt the dog just for the pleasure of doing so. Do you say, "That was a vicious act," or "That was a vicious act on condition that it issued from a stable disposition to give similar kicks in similar circumstances"? Surely you say the former. Or imagine that your companion stops to give \$20 to a homeless person, apparently from concern for that person for her own sake. Do you say "That was generous of you," or "That was generous of you on condition that it issued from a stable disposition to act from similar motives in similar circumstances"? Again, surely you say the former. Since your judgment is of virtuousness, it turns on your companion's motives rather than on any external features of her act. If you learn that she was acting only to impress you or some bystanders, you may grant that she acted rightly but will withdraw your attribution of generosity.⁶ But that attribution concerns only her current motives, apart from any connection to longer-lasting traits. Or imagine that a military committee is considering whether to give a soldier a medal for bravery. Would they say, "We know he threw himself on a grenade despite knowing it would cost him his life and in order to save the lives of his comrades. But we cannot give him a medal for bravery because we do not know whether his act issued from a stable disposition or was, on the contrary, out of character"? They would say no such thing, and they would be obnoxious if they did.

A defender of the dispositional view may reply that these points are not decisive. Everyday thought can locate virtue primarily in character traits but think that in each of the three examples the act described is sufficient evidence for a trait. If

³ Gary Watson, "On the Primacy of Character," in *Identity, Character and Morality: Essays in Moral Psychology*, ed. O. Flanagan and A. O. Rorty (Cambridge, Mass.: MIT Press, 1990), pp. 455, 459; Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Clarendon Press, 1999), pp. 29, 134–36, 157–59, 167.

⁴ Roger Crisp and Michael Slote, Introduction, in *Virtue Ethics*, ed. Crisp and Slote (Oxford: Oxford University Press, 1997), p. 3.

⁵ Julia Annas, "Virtue Ethics," in *Oxford Handbook of Ethical Theory*, ed. David Copp (Oxford: Clarendon Press, 2005), p. 534 n. 3.

⁶ Some virtue terms such as "just" have a different use that is independent of the agent's motives. Thus we may say that a storekeeper who gives accurate change performs a just act even if he does so only to avoid losing customers; here "just" indicates a ground of rightness rather than anything connected to motivation. But other virtue-terms have no such use. To call an act brave, generous, or kind is always to say something about the agent's motives.

this were so, however, everyday thought would recognize the relevance of facts about a person's behavior at other times to the question whether his current act is virtuous, and it does not. It is even happy to call out-of-character acts virtuous. If the companion who now shows concern for the homeless person has never done so before, you may say, "That was uncharacteristically generous of you"; if you do, you will not contradict yourself.

If the issue between the dispositional and occurrent-state views concerned only the use of "virtue" and related terms, it would not be of great philosophical significance. But it is usually connected to a moral issue. However exactly it is understood, the concept of virtue is that of a state that is somehow desirable; let us assume it is the concept of a state that is good in itself. Then when philosophers disagree about the primary application of the term "virtue" it is usually because they disagree about what in this area is primarily good: the one side says it is dispositions, the other says it is occurrent states apart from dispositions. We need to address this issue, but should first note some slightly different views each terminological claim can express.

For an extreme dispositional view, consider the neo-Aristotelian formula that identifies the virtues as those traits a person needs in order to flourish or live well. Read literally, this formula says that what contributes to a person's flourishing is only his virtuous traits as traits; though having them will lead him to act virtuously, his virtuous acts are mere side effects of what has value but have no worth in themselves. Aristotle's own view was less extreme. While defining virtue initially as a disposition, he recognized that it can be possessed by someone "who is asleep or in some other way quite inactive"; on that basis he held that the prime contributor to flourishing is the active exercise of virtue, found in occurrent virtuous acts, desires, and feelings.⁷ But since he counted these states as virtuous only when they issue from a virtuous disposition, he continued to give priority to dispositions as, if not the prime bearers of value, then necessary conditions for what are. A dispositional view can even define the virtues as dispositions to act from the motives deemed virtuous by Ross, such as a desire for another's pleasure for its own sake. But by calling acts virtuous only when they issue from a stable trait, it still places evaluative priority on dispositions as at least necessary conditions for what is primarily good.

In contrast, by applying the term "virtue" to acts and feelings regardless of their connection to stable traits, the occurrent-state view finds the primary value in these states considered on their own. This is of course consistent with finding various kinds of value in virtuous dispositions. Most obviously, such dispositions can have great instrumental value, since they tend to produce individual virtuous acts and feelings and, through them, further benefits such as pleasures for other people. Virtuous dispositions can even be the prime source of virtuous acts or the most

⁷ *Nicomachean Ethics*, 1095b32–33, 1098b32–1099a6.

reliable means of producing them, so moral education should make their development its central goal. In addition, the occurrent-state view can hold that virtuous dispositions have some intrinsic value as dispositions. Ross took this line, saying the state of mind of a habitually unselfish person is intrinsically better than that of a habitually selfish one even when neither is exercising his disposition.⁸ But this intrinsic value will typically be less than that of the occurrent virtuous states the dispositions issue in, and the view will continue to emphasize such states by making their value independent of any connection to longer-lasting traits.

I have said that everyday moral thought accepts the occurrent-state rather than the dispositional view of virtue. Given the moral claims usually implicit in the two views, I think it is right to do so. The question is whether an act performed from a given occurrent motive, such as a desire for another's pleasure for its own sake, is less good in itself if it does not issue from a stable disposition. I see no reason to believe this.

The presence of such a disposition will typically lead a person to act virtuously on many other occasions, making his life as a whole much better than if such acts were only occasional. The disposition may also have, as Ross held, some value in itself. But if we are to assess the moral claims of the dispositional and occurrent-state views we must abstract from these facts. We must imagine two acts with the same occurrent motive—say, the same desire for another's pleasure for its own sake—with the same motivational force, but where one desire issues from a stable trait of character and the other does not. Then, ignoring any values in other states associated with the two acts, we must ask whether the first act is in itself better, or more deserving of praise, than the second. I see no reason whatever to believe this. An act of helping another from genuine concern for his welfare is no less admirable if it happens to be out of character, and self-sacrifice in battle not an iota less deserving of a medal.

A defender of the dispositional view may reject the above scenario, where an in-character and an out-of-character act have the same occurrent motivation, as unrealistic. Someone who, say, helps a homeless person when he has not done so before is likely to be moved by some trivial feature of the situation, such as the particular expression on the person's face, or be to be acting on a whim that would not issue in action given even a moderately strong contrary impulse. Either way his desire will not have the depth or motivational force needed for it to have significant value. But if he acts from a stable disposition his desire will have those features, and it is this connection between stability and depth that makes character traits the prime locus for the value of virtue.⁹

⁸ *Foundations of Ethics* (Oxford: Clarendon Press, 1939), pp. 291–92.

⁹ Hursthouse, *On Virtue Ethics*, pp. 134–36, 157–59. Similar points are made, though not in aid of the dispositional view, in Nomy Arpaly, *Unprincipled Virtue* (New York: Oxford University Press, 2003), pp. 94–97.

There are two answers to this objection. First, while the connection it posits between stability and depth of motivation often obtains, it does not do so always. Someone can be stably disposed to act from a motive that is quite weak, so long as the only motives that ever oppose it are weaker. And an out-of-character act can be both focused on what is morally central and motivationally powerful; a soldier who has previously been timorous can now want to save his comrades just because they are his comrades, and can care deeply enough about doing so that he sacrifices his life. Second, even granting the connection, the objection gives no support to the dispositional view. It treats the existence of a disposition only as evidence for properties of depth and strength that belong to occurrent states as occurrent; it is at the time he acts that, for example, the soldier's desire to save his comrades is stronger than his desire to save himself. The objection therefore concedes that the primary intrinsic values are found in occurrent states and abandons any intrinsic concern for traits of character.¹⁰

The dispositional view was Aristotle's, and it is also common among contemporary virtue ethicists. These two points are connected, since much contemporary virtue ethics is strongly influenced by Aristotle. But while it is often valuable to study classical philosophical texts, in this case too much attention to ancient philosophy can blind one to what I think are obvious facts about the everyday understanding of virtue. Commonsense morality certainly makes global judgments about virtue. It can say that a given person is brave or has the standing trait of generosity. But it treats those judgments as derivative from local judgments about the virtuousness of particular acts, desires, and feelings, and takes those states' virtuousness to be independent of any tie to dispositions. Moreover, it is right to do so: an act of helping another from a desire for his welfare is no less admirable when out of character than when dispositionally based.

I will close with a final comment. There has been much discussion recently of the philosophical implications of situationist social psychology, which denies that most people have stable traits of character and says that much of their behavior is influenced by trivial-seeming features of their situations. Philosophers impressed by situationism have said it challenges both folk psychology, which explains actions by reference to stable dispositions, and the part of commonsense morality that is focused on virtue and vice. Of these challenges, the one to commonsense morality would have considerable force if that morality took the dispositional view of virtue. If it were a necessary condition for virtuous action that it issue from a stable trait of character, and few people had such traits, then common sense would be wrong to apply the virtue terms to most of the acts people perform. (Virtuous action could still be a moral ideal, just one that few people ever achieve.) But the challenge

¹⁰ Of course one can define "depth" of motivation so it requires persistence through time; Hursthouse sometimes seems to do this. But then the objection's defense of the dispositional view begs the question, using a concept of "depth" that no one moved by the examples I have appealed to will think reflects any value.

has no force if, as I have argued, commonsense morality accepts the occurrent-state view of virtue. Then if most people will help another from concern for her happiness in situation *A* but not in trivially different situation *B*, situation *A* causally encourages their acting generously while situation *B* does not. Knowing this fact will be important practically. If we want to promote generous action, we should place people in situations of type *A* rather than of type *B*. But the situationist fact will have no effect on the concept of virtuous action, which is just that of action from an occurrent virtuous desire no matter what that desire's causes are. Some writers on situationism have recognized this point, saying the theory poses no challenge to a morality of virtue focused on particular acts and feelings rather than on traits of character.¹¹ But they have not connected this point to commonsense morality, which they have tended to assume accepts the dispositional view and therefore is open to situationist challenge. Perhaps they have been overly influenced by neo-Aristotelian writing about virtue; perhaps they have assumed that if common sense takes dispositions to be central to the psychological explanation of particular acts, it must also take them to be central to their evaluation. But there is no reason to expect this last connection, and commonsense morality in fact accepts the occurrent-state view. It is therefore not in the least threatened by situationist social psychology, but can happily take it on board.¹²

¹¹ Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99 (1999): 327–28; John Doris, *Lack of Character: Personality and Moral Behaviour* (Cambridge: Cambridge University Press, 2002), pp. 116–117.

¹² I thank Danielle Bromwich for research assistance, and Donald Ainslie, Phil Clark, Gopal Sreenivasan, and Sergio Tenenbaum for helpful comments on an earlier draft.

Games and The Good

Our societies attach considerable value to excellence in sports. In Canada hockey players are named to the highest level of the Order of Canada; in Britain footballers and cricketers are made MBE and even knighted. And this attitude extends more widely. Sports are a subclass of the wider category of games, and we similarly admire those who excel in nonathletic games such as chess, bridge, and even Scrabble.

I take this admiration to rest on the judgment that excellence in games is good in itself, apart from any pleasure it may give the player or other people, but just for the properties that make it excellent. The admiration, in other words, rests on the perfectionist judgment that skill in games is worth pursuing for its own sake and can add value to one's life. This skill is not the only thing we value in this way; we give similar honours to achievements in the arts, science and business. But one thing we admire, and to a significant degree, is excellence in athletic and nonathletic games.

Unless we dismiss this view, one task for philosophy is to explain why such excellence is good. But few philosophers have attempted this, for a well-known reason. A unified explanation of why excellence in games is good requires a unified account of what games are, and many doubt that this is possible. After all, Wittgenstein famously gave the concept of a game as his primary example of one for which necessary and sufficient conditions cannot be given but whose instances are linked only by looser "family resemblances."¹ If Wittgenstein was right about this, there can be no single explanation of why skill in games is good, just a series of distinct explanations of the value of skill in hockey, skill in chess, and so on.

But Wittgenstein was not right, as is shown in a little-known book that is nonetheless a classic of twentieth-century philosophy, Bernard Suits's *The Grasshopper*:

¹ Ludwig Wittgenstein, *Philosophical Investigations*, 3rd ed., trans. G. E. M. Anscombe (Oxford: Blackwell, 1972), Sect. 66.

Games, Life and Utopia. Suits gives a perfectly persuasive analysis of playing a game as, to quote his summary statement, “the voluntary attempt to overcome unnecessary obstacles.”² And in this paper I will use his analysis to explain the value of playing games. More specifically, I will argue that the different elements of Suits’s analysis give game-playing two distinct but related grounds of value, so it instantiates two related intrinsic goods. I will also argue that game-playing is an important intrinsic good, which gives the clearest possible expression of what can be called a modern as against a classical—or more specifically, Aristotelian—view of value.

But first Suits’s analysis. It says that a game has three main elements, which he calls the prelusory goal, the constitutive rules, and the lusory attitude. To begin with the first, in playing a game one always aims at a goal that can be described independently of the game. In golf, this is that a ball enter a hole in the ground; in mountain climbing, that one stand on top of a mountain; in Olympic sprinting, that one cross a line on the track before one’s competitors. Suits calls this goal “prelusory” because it can be understood and achieved apart from the game, and he argues that every game has such a goal. Of course, in playing a game one also aims at a goal internal to it, such as winning the race, climbing the mountain, or breaking par on the golf course. But on Suits’s view this “lusory” goal is derivative, since achieving it involves achieving the prior prelusory goal in a specified way.

This way is identified by the second element, the game’s constitutive rules. According to Suits, the function of these rules is to forbid the most efficient means to the prelusory goal. Thus, in golf one may not carry the ball down the fairway and drop it in the hole by hand; one must advance it using clubs, play it where it lies, and so on. In mountain climbing one may not ride a gondola to the top of the mountain or charter a helicopter; in 200-metre sprinting, one may not cut across the infield. Once these rules are in place, success in the game typically requires achieving the prelusory goal as efficiently as they allow, such as getting the ball into the hole in the fewest possible strokes or choosing the best way up the mountain. But this is efficiency within the rules, whose larger function is to forbid the easiest means to the game’s initial goal.

These first two elements involve pursuing a goal by less than the most efficient means, but they are not sufficient for playing a game. This is because someone can be forced to use these means by circumstances he regrets and wishes were different. If this is the case—if, for example, a farmer harvests his field by hand because he cannot afford the mechanical harvester he would much rather use—he is not playing a game. Hence the need for the third element in Suits’s analysis, the lusory attitude, which involves a person’s willingly accepting the constitutive rules, or accepting them because they make the game possible. Thus, a golfer accepts that

² Bernard Suits, *The Grasshopper: Games, Life and Utopia* (Toronto: University of Toronto Press, 1978; repr. Peterborough, ON: Broadview Press, 2005), p. 41/55 (page references are first to the University of Toronto Press edition, then to the Broadview Press edition).

he may not carry the ball by hand or improve his lie because he wants to play golf, and obeying those rules is necessary for him to do so; the mountaineer accepts that he may not take a helicopter to the summit because he wants to climb. The restrictions the rules impose are adhered to not reluctantly but willingly, because they are essential to the game. Adding this third element gives Suits's full definition: "To play a game is to attempt to achieve a specific state of affairs [prelusory goal], using only means permitted by the rules...where the rules prohibit the use of more efficient in favor of less efficient means [constitutive rules], and where the rules are accepted just because they make possible such activity [lusory attitude]." Or, in the summary statement quoted above, "playing a game is the voluntary attempt to overcome unnecessary obstacles."³

This analysis will doubtless meet with objections, in the form of attempted counterexamples. But Suits considers a whole series of these in his book, showing repeatedly that his analysis handles them correctly, and not by some ad hoc addition but once its elements are properly understood. Nor would it matter terribly if there were a few counterexamples. Some minor lack of fit between his analysis and the English use of "game" would not be important if the analysis picks out a phenomenon that is unified, close to what is meant by "game," and philosophically interesting. But the analysis is interesting if, as I will now argue, it allows a persuasive explanation of the value of excellence in games.

Suits himself addresses this issue of value. In fact, a central aim of his book is to give a defense of the grasshopper in Aesop's fable, who played all summer, against the ant, who worked. But in doing so he argues for the strong thesis that playing games is not just an intrinsic good but the supreme such good, since in the ideal conditions of utopia, where all instrumental goods are provided, it would be everyone's primary pursuit. The grasshopper's game-playing, therefore, while it had the unfortunate effect of leaving him without food for the winter, involved him in the intrinsically finest activity. Now, I do not accept Suits's strong thesis that game-playing is the supreme good—I think many other states and activities have comparable value—and I do not find his arguments for it persuasive. But I will connect the weaker thesis that playing games is one intrinsic good to the details of his analysis more explicitly than he ever does.

Consider the first two elements of the analysis, the prelusory goal and constitutive rules. By forbidding the most efficient means to that goal, the constitutive rules usually make for an activity that is reasonably difficult. They do not always do so. Rock, paper, scissors is a game whose prelusory goal is to throw rock to one's opponent's scissors, scissors to his paper, or paper to his rock, and the rules forbid the easiest means to this goal by forbidding one to make one's throw after he has made his. But though the rules make achieving this goal more difficult than it might be, they do not make it by absolute standards difficult; rock, paper, scissors is not a challenging activity. But then rock, paper,

³ Suits, *The Grasshopper*, p. 41/54–5.

scissors is not a very good game, and certainly not one the playing of which has much intrinsic value. It is characteristic of good games to be not only more difficult than they might be but also in absolute terms reasonably difficult. They cannot be so difficult that no one can succeed at them, but also cannot lack all challenge; they must strike a balance between too much and too little difficulty. In what follows I will defend the value only of playing good games, because they realize what seems an internal goal of the design of games. If the constitutive rules of a game make achieving its prelusory goal more difficult than it might be, this is surely because they aim at making it simply difficult.

If the prelusory goal and rules of a good game make succeeding at it reasonably difficult, they will also give it one ground of value if difficult activities are as such intrinsically good. And I believe that difficult activities are as such good. Though not often explicitly affirmed by philosophers, this view can be defended in at least two ways.

Many contemporary philosophers include among their intrinsic goods achievement, by which they mean not just moral but also nonmoral achievement—for example, in business or the arts.⁴ But what exactly is achievement? It clearly involves realizing a goal, but not every such realization counts as an achievement; for example, tying one's shoelace does not unless one has some disability. And among achievements some are more valuable than others; thus, starting a new business and making it successful is a greater achievement than making a single sale. If we ask what explains these differences—between achievements and nonachievements, and between greater and lesser achievements—the answer is surely in large part their difficulty: how complex or physically challenging they are, or how much skill and ingenuity they require. It is when a goal is hard to bring about that doing so is an achievement. So reflection on our intuitive understanding of the value of achievement suggests a first reason for holding that difficult activities are as such good.

A second reason, which is complementary but more abstract, is suggested by Robert Nozick's fantasy of an "experience machine."⁵ This machine, which can electrically stimulate the brain to give one the pleasure of any activity one wants, is intended as a counterexample to the hedonistic view that only pleasure is good, but it also makes a positive point. If life on the machine is less than ideal, this is largely because people on it are disconnected from reality. They have only false beliefs about their environment and never actually realize any goals: they may think they are discovering a cure for cancer or climbing Everest, but in fact they are not. This suggests that an important good is what we can call "rational connection to reality," where this has two aspects, one theoretical and one practical.⁶

The theoretical aspect is knowledge, or having beliefs about the world that are both true and justified. The beliefs' truth means there is a match between one's mind and

⁴ See, for example, James Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Clarendon Press, 1986), p. 67.

⁵ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 42–5.

⁶ I give a fuller account of this value in my *Perfectionism* (New York: Oxford University Press, 1993), chs. 8–10.

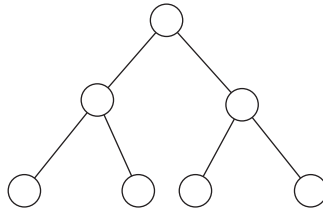


FIGURE 11.1 *Structured Knowledge*

reality; their being justified means the match is not a matter of luck but something one’s evidence made likely. But a full account of this good must explain which kinds of knowledge are most worth having. Classical philosophers like Aristotle thought the best knowledge is of the intrinsically best objects, such as the divine substances, but the more plausible view is that the best knowledge has the most of certain formal properties that are independent of its subject matter. More specifically, the best knowledge is explanatorily integrated, with general principles that explain middle-level principles that in turn explain particular facts. This integration results in an explanatory hierarchy like that represented in Figure 11.1, where items of knowledge higher up in the hierarchy explain those below them. And this hierarchy embodies more intrinsic value than if one knew only isolated unexplanatory facts, like the number of grains of sand on seven beaches (Figure 11.2). We can give an artificial but illustrative model for measuring this value if we imagine that each item of knowledge initially has one unit of value in itself, but gains an extra unit for every other item of knowledge subordinate to it in a hierarchy. Then the seven isolated items in Figure 11.2 have just one unit of value each, for a total of seven units. But in Figure 11.1 the middle items have three units, since they each explain two further facts, and the top item has seven units, for a total of seventeen units in the hierarchy as a whole. The explanatory relations between them give an integrated set of beliefs more value than ones that are unconnected.

This model can be enriched. We may think it especially valuable to give unifying explanations of diverse facts, or to make surprising connections between what seemed unrelated phenomena. If so, we can count not just the number of individual items a given item of knowledge has subordinate to it, but the number of items of different kinds, so there is more value in explaining more types of fact. We may also value precision of knowledge, such as knowing that the constant of gravitational acceleration is not just somewhere between 5 and 15 m/s² but exactly 9.8m/s². And we can capture this view both by giving more value to precise knowledge in itself and by giving it more additional value for explaining further precise truths.

Finally, we may think that knowing truths concerning many objects is better than knowing highly particular ones, even apart from the former’s



FIGURE 11.2 *Unstructured Knowledge*

explanatory role; thus, knowing a scientific law is better than knowing the number of grains of sand on some beach even if one has not used the former to explain anything else.

The practical parallel to knowledge, and the other value missing on the experience machine, is achievement, or realizing a goal in the world given a justified belief that one would do so. Here again there is a match between one's mind and reality, though now reality has been made to fit one's mind, and a justified belief that makes the match not just lucky. Again we must specify which achievements are best. A classical view might say they are of the goals that are independently best, but we can maintain the parallel with knowledge, and give a better account of achievement as achievement, if we say they are of the goals with the most of certain formal properties that again center on hierarchical integration. This time, however, the integrating relation is not explanatory but means-end. Thus, in Figure 11.1 we achieve the goal at the top of the hierarchy by achieving the two middle-level goals as means to it, and each of those by achieving the two below them. And if each nonluckily achieved goal has one unit of value in itself plus an additional unit for every other goal achieved as a means to it, the achievements in this hierarchy again have seventeen units of value as against the seven in seven unrelated achievements. Just as more complex explanatory relations make for more value in knowledge, so more complex means-end relations make for more value in achievement.

Again this model can be enriched. We may think achievements are especially valuable if they require subsidiary achievements of varied kinds, and can capture this view by counting the number of goals of different types a given one has subordinate to it. More strongly, we may deny significant value to achievements that involve only subordinate goals of the same repetitive type. We may also value precision in achievement—hitting a particular target rather than just some vague area—and can give achievements additional value for that. And we can think that, apart from means-end relations, achieving goals whose content extends more widely, through time or in the number of objects they involve, is likewise more valuable.

This model deepens the value of achievement by showing it to be parallel to knowledge and, with it, one aspect of a more abstract good of rational connection to reality. It also makes many difficult activities good for the very properties that make them difficult. First, the more complex the means-end hierarchy an activity involves, the more places there are where one can fail at something crucial and the harder success in the activity becomes. Second, the more complex the hierarchy, the more deliberative skill it requires, since one has to monitor one's progress through a more elaborate sequence of tasks. There is a further increment of difficulty if the hierarchy involves a greater variety of subordinate goals, since then it requires a greater variety of skills, and likewise if the activity demands more precision. And it is more difficult to achieve goals with more extended contents, both

because holding them in one's mind is more difficult and because achieving them requires changing more of the world.⁷

Moreover, these are precisely the aspects of difficulty found in good games. These games usually require one to go through a complex sequence of tasks rather than do one simple thing such as throw rock, paper, or scissors. The tasks in question often demand varied skills: thus, golf requires one not only to drive the ball a long distance but to drive it accurately, play from bunkers, putt, and make strategic decisions. Good golfers are also precise, hitting their approach shots to a particular part of the green rather than just somewhere near it. And many games, such as chess, hockey, and basketball, require players to grasp an extended content, including all the pieces on the board or all the players on the ice or court, in a single act of consciousness. That again is difficult, and requires years of practice to master.

Not all the difficulty in games involves this complex ratiocination. Weightlifters have to go through a precisely ordered sequence of moves in order to lift their weights, but also need brute strength: if one of two lifters has less perfect technique but is stronger and therefore lifts more, he wins the competition. Boxing, too, depends in part on raw power. These purely physical forms of difficulty do not instantiate the value of rational connection, and their role in making game-playing good is unclear. Why do we value the physical aspects of weightlifting and boxing but not those found in, say, pie-eating contests? Does this reflect just the historical accident that weightlifting and boxing began long enough ago that we can value them now for their traditions? Or do we value physical difficulty only when it accompanies more rational forms of challenge but not on its own? I will not pursue this issue, taking the rational connection model to capture what makes purely cerebral games such as chess difficult, and also much of what makes sports such as golf and hockey difficult.

I have argued that the prelusory goal and constitutive rules make playing a good game difficult, and have given two reasons to believe that difficulty is as such good. But I have not yet used the third element in Suits's analysis, the lusory attitude. Let us examine it more closely.

In his 1907 book *The Theory of Good and Evil*, Hastings Rashdall remarked that "sport has been well defined as the overcoming of difficulties simply for the sake of overcoming them."⁸ This definition is close to Suits's, but differs on one point. It in effect takes the lusory attitude to be one of accepting the rules because they make the game difficult, whereas Suits takes it to be one of accepting the rules because

⁷ Some may deny that difficulty is as such good, on the ground that an activity aimed at evil, such as genocide, is not in any way made good by its difficulty. The issue here is complex (see my *Virtue, Vice, and Value* [New York: Oxford University Press, 2001], pp. 144–52), but those moved by this objection can retreat to the weaker claim that only activities with good or neutral aims gain value by being difficult. This weaker claim is sufficient to ground the value of games.

⁸ *The Theory of Good and Evil*, 2 vols. (Oxford: Oxford University Press, 1907), vol. 2, p. 105.

they make the game possible. For Rashdall, the golfer accepts the rule against improving one's lie because it makes golf harder; for Suits, it is because it makes golf golf. Which view is correct?

Suits's view is preferable if we are analyzing the generic concept of playing a game. Consider what we can call a pure professional golfer, who plays golf only as a means to making money and with no interest in the game for itself. He does not cheat as a means to making money; he knows that to make money he must play golf, which means obeying all its rules. But his only reason for accepting the rules is to make money. If we used Rashdall's view to define the generic concept, we would have to say the pure professional is not playing golf, which is absurd. But on Suits's view he is playing golf: though he accepts the rules only as a means to money, he does accept them in order to play golf and so has the lusory attitude.

But though Suits defines the generic concept of game-playing, this is not what he defends as the supreme intrinsic good. His argument, recall, is that in utopia, where all instrumental goods are provided, game-playing would be everyone's primary activity. But this description of utopia implies that it would contain no professional players; since no one would need to play a game as a means to anything, all players would be amateurs who chose the game for itself. But then they would have Rashdall's lusory attitude of accepting the rules because they make the game difficult, and Suits explicitly agrees. He describes how one utopian character decides to build houses by carpentry rather than order them up telepathically because carpentry requires more skill. And he starts his discussion of utopia by saying he will defend the value of game-playing as a specific form of play, where he has earlier denied that playing a game necessarily involves playing: to play is to engage in an activity for its own sake, and a pure professional does not do that.⁹ So the activity Suits defends as supremely good is game-playing that is also play, or what I will call "playing in a game." And that activity involves accepting the rules not just because they make the game possible, but also because they make it difficult.

I will follow Suits here and narrow my thesis further: not only will I explain the value only of playing good games, I will explain the value only of playing *in* these games, or of playing them with an at least partly amateur attitude. But this is not in practice much of a restriction, since most people do play games at least partly for their own sakes. Consider Pete Rose, an extremely hard-nosed baseball player who was disliked for how much he would do to win. Taking the field near the end of the famous sixth game of the 1975 World Series, and excited by the superb plays that game had involved, he told the opposing team's third base coach, "Win or lose, Popeye, we're in the fuckin' greatest game ever played"; after the game, which his

⁹ Suits, *The Grasshopper*, pp. 166/149, 144/130.

team lost, he made a similar comment about it to his manager. Intensely as he wanted to win, Pete Rose also loved baseball for itself.¹⁰

So the game-playing whose value I will explain involves accepting the rules of the game because they make it difficult. But then the elements that define this type of game-playing are internally related: the prelusory goal and constitutive rules together give it a feature, namely difficulty, and the lusory attitude chooses it because of this feature. More specifically, if difficulty is as such good, the prelusory goal and rules give it a good-making feature and the lusory attitude chooses it because of that good-making feature. This connects the lusory attitude to an attractive view that has been held by many philosophers, namely that if something is intrinsically good, the positive attitude of loving it for the property that makes it good—that is, desiring, pursuing, and taking pleasure in it for that property—is also, and separately, intrinsically good. Thus, if another person's happiness is good, then desiring, pursuing, and being pleased by her happiness as happiness is a further good, namely that of benevolence; likewise, if knowledge is good, then desiring, pursuing, and being pleased by knowledge is good. Aristotle expressed this view when he said that if an activity is good, pleasure in it is good, whereas if an activity is bad, pleasure in it is bad,¹¹ and it was accepted around the turn of the twentieth century by many philosophers, including Rashdall, Franz Brentano, G. E. Moore, and W. D. Ross. And it applies directly to playing in games, which combines the good of difficulty with the further good of loving difficulty for itself. The prelusory goal and constitutive rules together give playing in games one ground of value, namely difficulty; the lusory attitude in its amateur form adds a related but distinct ground of value, namely loving something good for the property that makes it so. The second ground depends on the first; loving difficulty would not be good unless difficulty were good. But it adds a further, complementary intrinsic good. When you play a game for its own sake you do something good and do it from a motive that fixes on its good-making property.

This two-part explanation deepens Suits's claim that playing in games is an intrinsic good, by connecting it to more general principles of value with application beyond the case of games. At the same time, however, it makes playing in games a derivative rather than a fundamental intrinsic good. It would not appear on a list of basic goods, since it combines two other, more fundamental, goods in a particular way.

But a good that is not fundamental can nonetheless be paradigmatic because it gives the clearest possible expression of a certain type of value. If difficult activities are as such good, they must aim at a goal: it is achieving that which is challenging.

¹⁰ Tom Adelman, *The Long Ball: The Summer of 75—Spaceman, Catfish, Charlie Hustle, and the Greatest World Series Ever Played* (New York: Back Bay Books, 2003), p. 313.

¹¹ Aristotle, *Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984), 1175b24–30. I discuss this view at length in *Virtue, Vice, and Value*.

But their value does not derive from properties of that goal considered in itself, depending instead on features of the process of achieving it. Yet this can be obscured if the goal is independently good, since then the activity, if successful, will be instrumentally good, and this can seem the most important thing about it. If the farmer who works by hand successfully harvests a crop, his work contributes to the vital good of feeding his family, and this can distract us from the value it has in itself. But there is no such danger if the goal is intrinsically valueless, as it most clearly is in games. Since a game's prelusory goal—getting a ball into a hole in the ground or standing atop a mountain—is intrinsically trivial, the value of playing the game can depend only on facts about the process of achieving that goal. And this point is further emphasized by the lusory attitude, which chooses that process just as a process, since it willingly accepts rules that make achieving the goal harder. Game-playing must have some external goal one aims at, but the specific features of this goal are irrelevant to the activity's value, which is entirely one of process rather than product, journey rather than destination. This is why playing in games gives the clearest expression of a modern as against an Aristotelian view of value: because modern values are precisely ones of process or journey rather than of the end-state they lead to.

The contrary Aristotelian view, which denigrates these values, was expressed most clearly in Aristotle's division of all activities into the two categories of *kinēsis* and *energeia*, and his subsequent judgments about them.¹² An Aristotelian *kinēsis*—often translated as “movement”—is an activity aimed at a goal external to it, as driving to Toronto is aimed at being in Toronto. It is therefore brought to an end by the achievement of that goal, which means that a *kinēsis* can be identified by a grammatical test: if the fact that one has *X*-ed implies that one is no longer *X*-ing, as the fact that one has driven to Toronto implies that one is no longer driving there, then *X*-ing is a *kinēsis*. But the main point is that a *kinēsis* aims at an end-state separate from it. By contrast, an *energeia*—translated variously as “actuality,” “activity,” or “action”—is not directed at an external goal but has its end internal to it. Contemplation is an *energeia*, because it does not aim to produce anything beyond itself, as is the state of feeling pleased. And *energeiai* do not pass the above grammatical test, and therefore, unlike *kinēseis*, can be carried on indefinitely: that one has contemplated does not imply that one is not contemplating now or will not continue to do so. Contemplation, like driving to Toronto, is an activity, but it does not aim to produce anything apart from itself.

Now, Aristotle held that *energeiai* are more valuable than *kinēseis*, so the best human activities must be ones that can be carried on continuously, such as contemplation. This is because he assumed that the value of a *kinēsis* must derive from that of its goal, so its value is subordinate, and even just instrumental, to that of the goal. As he said at the start of the *Nicomachean Ethics*,

¹² *Nicomachean Ethics*, 1094a1–7, 1174a13–b8, 1176b1–8, 1177b2–4.

“Where there are ends apart from the actions, it is the nature of the products to be better than the activities.”¹³ But it is characteristic of what I am calling modern values to deny this assumption, and to hold that there are activities that necessarily aim at an external goal but whose value is internal to them in the sense that it depends entirely on features of the process of achieving that goal. Suits cites expressions of this modern view by Kierkegaard, Kant, Schiller, and Georg Simmel,¹⁴ but for an especially clear one consider Marx’s view that a central human good is transforming nature through productive labor. This activity necessarily has an external goal—one cannot produce without producing some thing—and in conditions of scarcity this goal will be something vital for humans’ survival or comfort. But Marx held that when scarcity is overcome and humans enter the “realm of freedom” they will still have work as their “prime want,” so they will engage in the process of production for its own sake without any interest in its goal as such. Or consider Nietzsche’s account of human greatness. In an early work he said the one thing “needful” is to “give style to one’s character,” so its elements are unified by “a single taste,” and that it matters less whether this taste is good or bad than whether it is a single taste.¹⁵ Later he said the will to power involves not the “multitude and disgregation” of one’s impulses but their coordination under a single predominant impulse.¹⁶ In both discussions he deemed activities good if they involve organizing one’s aims around a single goal whatever that goal is. So for both Marx and Nietzsche a central human good was activity that on the one side is necessarily directed to a goal but on the other derives its value entirely from aspects of the process of achieving it. This is why the type of value they affirm is paradigmatically illustrated by playing in games; when one’s goal is trivial, the only value can be that of process. Marx and Nietzsche would never put it this way, but what each valued is in effect playing in games: in Marx’s case the game of material production when there is no longer any instrumental need for it, in Nietzsche’s the game of exercising power just for the sake of doing so.

Playing in games also clearly straddles Aristotle’s division between *kinēseis* and *energeiai*. It has the logical structure of a *kinēsis*, since it aims at a goal external to itself, and passes the relevant grammatical test: if one has parred a golf hole or climbed a mountain, one is no longer doing so. But it also has value in itself, as an *energeia* does, based on properties internal to it as an activity. We can show this more precisely using our formal model of the value of achievement, on which the value of any goal depends in part on the number of other goals achieved as means to it. In Figure 11.1 the lower-level goals are pursued as means to higher-level ones,

¹³ *Nicomachean Ethics*, 1094^a14–5.

¹⁴ Suits, *The Grasshopper*, pp. 93–94/92.

¹⁵ Friedrich Nietzsche, *The Gay Science*, trans. Walter Kaufmann (New York: Vintage, 1974), sec. 290.

¹⁶ Nietzsche, *The Will to Power*, trans. Walter Kaufmann and R. J. Hollingdale (New York: Vintage, 1968), sec. 46.

and contribute to those goals' value only if they are both successfully achieved and contribute causally to them. And the higher-level goals must themselves also be successfully achieved. Since the hierarchy is precisely one of *achievements*, a highest-level goal that is not achieved does not qualify for inclusion in the hierarchy, and so does not gain any value from having other goals achieved as attempted means to it. This means that if two people go through the same complex process as a means to a given goal, and the first achieves the goal while the second through bad luck does not, the first's activity has more intrinsic value: his hierarchy contains his highest-level goal, which has his greatest value, but the second's does not. (If Pete Rose's opponents played as well as he did but Rose's team won the World Series, his play was intrinsically better.) So the activities valued by our formal model are directed at an external goal, as *kinēseis* are, and have their full value only if that goal is achieved. But their value does not depend on properties of the goal considered by itself; if the same goal were achieved without complex means, it might have just one unit of value. Instead, their value depends on means–end relations between their components, and so depends on internal features of the activity as does that of an *energeia*.

If playing in games is the paradigm expression of modern values, it helps us see similar value in other activities not normally associated with games. One, emphasized by Nietzsche, is a life organized around a single goal; it embodies through a longer stretch of time the same hierarchical structure present in individual difficult activities. The relevant activities also include ones in business and the arts. Business activity sometimes aims at an independent good such as relieving others' suffering or increasing their comfort. But often its goal is just to win market share and profits for one company, which is morally trivial; there is no intrinsic value in people's drinking Coke rather than Pepsi or using Microsoft rather than Apple. Aristotle should therefore deny this activity value, and he did, arguing that if money has no intrinsic value, the activity of moneymaking must likewise have no value.¹⁷ But if winning market share is difficult, requiring a complex series of finely balanced decisions, a modern view can grant it significant worth. And its pursuit can also involve something like the lusus attitude, since business people who aim partly for profits can also value the exercise of business skill just as skill, or for its own sake. Artistic creation too, to cite a different activity, has an independently good product if it aims, say, at communicating truths that cannot be communicated by nonartistic means. But a distinctively modern view (which is not to say the only view held nowadays) says that art aims only at beauty, where that consists

¹⁷ *Nicomachean Ethics*, 1096a5–10. An obvious suggestion is that an activity like moneymaking can be a *kinēsis* when described in one way and an *energeia* when described in another. But, plausible though it is, this does not seem to have been Aristotle's view. He seems to have treated the distinction as a metaphysical one, between types of activities as they are in themselves. Nor could he have accepted the suggestion and continued to give his arguments about the inferiority of money-making and the superiority of contemplation, however described, based on their properties as *kinēsis* or *energeia*.

in organic unity, or having the different elements of a painting, novel, or piece of music form a coherent, dynamic whole. This view makes the value of artistic production rest on its intentionally creating all the complex relations that define its product's beauty—that is, on its itself being complex. And its value will be greater if it has more of the supplementary qualities mentioned above: if it unifies more varied elements; if it requires more precise brushstrokes, notes, or words; and if it involves grasping more extended contents in a single act of consciousness, as Henry Moore could see his sculptures from all sides at once.¹⁸ And of course artistic creation can involve a lusus attitude, if the artist enjoys and values the skill his work involves for its own sake.

But playing in games is also in one respect a lesser good, and I want to close by explaining why. Imagine two activities that are equally complex and difficult, one of which produces an intrinsically good result while the other does not. Perhaps one is political activity that liberates an entire nation from oppression while the other involves winning a high-level chess tournament. The first activity will, of course, be instrumentally better, because it produces a separate intrinsic good. But it will also arguably be on that basis intrinsically better. Consider Derek Parfit's example of a person who spends his life working for the preservation of Venice. Parfit claims, plausibly, that if after this person's death Venice is preserved, and in a way that depends crucially on his efforts, that will make his life and activities intrinsically better than if Venice had been destroyed.¹⁹ This conclusion already follows from our formal model of achievement, since any realization of a topmost goal adds value to a hierarchy. But I think there is an extra ground for its truth if, as Parfit clearly intends, the preservation of Venice is independently good. Whatever additional value there is in achieving a goal just as a goal, there is further value in achieving one that is good. When an activity aimed at a valuable end successfully achieves that end and therefore is instrumentally good, its being instrumentally good is an extra source of intrinsic goodness.²⁰

Now, because game-playing has a trivial end result, it cannot have the additional intrinsic value that derives from instrumental value. This implies that excellence in games, though admirable, is less so than success in equally challenging activities that produce a great good or prevent a great evil. This seems intuitively right: the honour due athletic achievements for themselves is less than that due the achievements of great political reformers or medical researchers. Whatever admiration we should feel for Tiger Woods or Gary Kasparov is less than we should feel for Nelson Mandela. It also implies that, whatever their other merits, Suits's utopia and Marx's realm of freedom would lack an important intrinsic good. Their

¹⁸ Howard Gardner, *Frames of Mind: The Theory of Multiple Intelligences* (New York: Basic Books, 1983), p. 188.

¹⁹ Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), p. 151.

²⁰ On this see Shelly Kagan, "Rethinking Intrinsic Value," *Journal of Ethics* 2 (1998): 277–97; and my "Two Kinds of Organic Unity," *Journal of Ethics* 2 (1998): 299–320.

inhabitants could play the game of, say, farming or medicine by going through the same complex procedures as farmers and doctors today. But if food could be produced and diseases cured by pushing a button, as they can in Suits's vision, their activity would not have the additional intrinsic value that comes from actually feeding or curing people and that is found in present-day farming and medicine.²¹ The very perfection of Suits's and Marx's utopias prevents them from containing the distinctive good of producing intrinsic goods that would not otherwise exist.

The point that an ideal world may exclude certain intrinsic goods should not be unfamiliar: G. E. Moore noted that the best possible world could not contain compassion for real pain, which he plausibly held was a greater good than compassion for merely imaginary pain.²² And Suits's and Marx's utopias can still contain, alongside such goods as pleasure and knowledge, the distinctively modern good of achieving a difficult goal regardless of its value. Moreover, their doing so can help make them better on balance than any world in which successful instrumental activity is possible. Many philosophers have assumed, with Aristotle, that the value of a process aimed at producing some end-state must derive entirely from the end-state's value, so if the latter is negligible so is the former. But there is no reason to believe this. Even if some of the process's intrinsic value depends on its instrumental value, in the way just described, there can also be intrinsic value in its properties just as a process and apart from any value in its product. To return again to Figure 11.1, this value will depend not on any qualities of the topmost goal considered in itself, but only on the means-ends relations between the various goals whose sequential achievement constitutes the process. I have argued that this distinctively modern value is illustrated most clearly by playing in games, especially when that is analyzed as in Bernard Suits's wonderful book *The Grasshopper*.²³

²¹ This claim is defended, with specific reference to Suits, in Shelly Kagan, "The Grasshopper, Aristotle, Bob Adams, and Me" (unpublished ms.).

²² G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), pp. 219–21.

²³ I am grateful for helpful conversations to my former student Gwendolyn Bradford, whose essay "Kudos for Ludus: Games and Value Theory" (*Noesis* 6 [2003]: 15–28) first linked the value of games and the details of Suits's definition of a game.

PART IV

Principles of Right

This page intentionally left blank

Rights and Capital Punishment

Discussions of the morality of capital punishment, and indeed discussions of the morality of punishment in general, usually assume that there are two possible justifications of punishment: a deterrence justification associated with utilitarianism and other consequentialist moral theories, and a retributive justification associated with deontological moral theories. But now that rights-based theories are attracting the increasing attention of moral philosophers it is worth asking whether these theories may not employ a different justification of punishment, with different consequences for the morality of particular forms of punishment. I will argue that rights theories do employ a different justification of punishment, and that this justification combines many of the attractive features of the deterrence and retributive justifications while avoiding their unattractive features. In particular, I will argue that the rights-based justification has more attractive consequences for the morality of capital punishment than either the deterrence or retributive justifications.¹

Rights-based moral theories hold that persons have certain natural rights, and the fact that these rights are natural is often expressed by saying that persons would possess them “in the state of nature.” Among the rights which persons are usually said to possess in the state of nature is the right to punish those who violate the rights of others. In section 7 of the *Second Treatise* Locke says that the state of nature has a Law of Nature to govern it, and that “every one has a right to punish

¹ Some philosophers have argued that we should apply the deterrence justification to the institution of punishment and the retributive justification to particular acts within this institution; for classic statements of this “mixed” view see John Rawls, “Two Concepts of Rules,” *Philosophical Review* 64 (1955): 3–32, and H. L. A. Hart, “Prolegomenon to the Principles of Punishment,” in his *Punishment and Responsibility* (Oxford: Clarendon Press, 1968), pp. 1–13. But these arguments seem to me to rely on a dubious distinction between an institution and the acts of which it is composed. The rights-based justification I will defend has many of the same attractive consequences as this mixed view without relying on its dubious assumptions about institutions.

the transgressors of the Law to such a Degree, as may hinder its Violation.”² Nozick too includes a right to punish among those he grants in *Anarchy, State, and Utopia*, quoting Locke’s description of this right with approval, and devoting an entire section to a discussion of “the right of all to punish.”³ If persons have a right to punish in the state of nature, then they are permitted to punish the violators of rights if they want to, but they are also permitted to refrain from punishing them if they do not want to. This already marks a certain departure from the deterrence and retributive views. Retributive theorists usually hold that we have a duty to punish, and that if we fail to punish when it is appropriate to do so we do something wrong. Deterrence theorists also hold that we have a duty to punish, and this means that the rights-based view, which holds that we never have more than a natural permission to punish, departs on at least this one point from the rigor of the two more familiar views. This departure is not, to be sure, one with any very important practical consequences. When persons move from the state of nature to civil society they transfer their right to punish to a government, and the transfer of this right is in fact definitive of the move to civil society. The transfer, however, is almost certainly conditional on the government’s undertaking to exercise its right whenever it is permissible for it to do so. The benefits which persons hope to gain by moving to civil society would be in serious jeopardy if their government could exercise its right to punish arbitrarily, and decide on grounds of its own choosing to punish some offenders and not others. This means that reasonable persons will only transfer their right to punish if their government agrees to exercise this right on every possible occasion, and that the practical implications of the rights-based justification are not significantly different on this point from justifications in which there is a positive duty to punish. Something like this certainly seems to have been the view of Locke. In section 219 of the *Second Treatise* he says that the social contract is dissolved whenever a government fails to exercise its right to punish, for a ruler who abandons his charge of executing the laws has done something the effect of which is “demonstratively to reduce all to Anarchy, and so effectively to *dissolve the Government*.”

The right to punish that persons have in the state of nature is not a primitive right, but derives from another more general right that they possess. Whenever persons in the state of nature have a natural right they also have the right to *enforce* that right—that is, the right to use coercion against other persons to prevent them from violating it.⁴ The most familiar form of coercion is the use of force, and persons in the state of nature therefore have the right to use force to defend themselves against would-be violators of their rights, and also to defend third parties. But this right of self- (and other-) defense is not the only enforcement right which

² John Locke, *Two Treatises of Government*, 2nd ed., ed. Peter Laslett (New York: Cambridge University Press, 1967).

³ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 10, 137–42.

⁴ On this see H. L. A. Hart, “Are There Any Natural Rights?” *Philosophical Review* 64 (1955): 175–91.

they possess. The making of threats is also a form of coercion, and persons in the state of nature therefore also have the right to threaten others with certain harms if they succeed in violating their rights, or succeed in violating the rights of third parties. It is from this second enforcement right that the right to punish derives. If persons in the state of nature have the right to threaten others with harms if they succeed in violating rights, then they surely also have the right to inflict these harms on them once the relevant rights have been violated. But this is just what the right to punish is: a right to inflict harms on persons who have successfully violated the rights of others.⁵

Although Locke and Nozick include a right to punish among those possessed in the state of nature they do not provide any justification of this right. They do not show *why* rights theories should contain a right to punish, or even why they should contain enforcement rights in general, but simply include these rights on a list of those possessed in the state of nature. There is one kind of rights theory, however, which in a somewhat stricter usage of the term than is usual I will call a “libertarian” rights theory, which can provide such a justification. A libertarian rights theory holds that there is really only one natural right, namely the equal right of all persons to the most extensive liberty compatible with a like liberty for other persons, and that all other natural rights are species or instances of the right to liberty. They are all rights to exercise liberty in certain specified areas, and impose on other persons the duty not to interfere with liberty in those areas.⁶ Because these rights are instances of a right to the *most extensive* liberty, we are to identify them by identifying the most extensive right to liberty possible. Comparing the extent of different liberties in the way this involves some obvious difficulties, but the following should be uncontroversial. If one liberty contains another as a proper part, so that exercising the second liberty always involves exercising the first, but exercising the first liberty does not always involve exercising the second, then the first liberty is more extensive than the second (some examples: the liberty to buy property in Canada is more extensive than the liberty to buy property in Alberta, for it contains it as a proper part; the liberty to move either of one’s arms freely is more extensive than the liberty to move one’s left arm freely, for it contains it as a

⁵ Enforcement rights are often said to include not only a right of self- (and other-) defense and a right to punish but also a right to exact compensation, where this right to exact compensation is sometimes exercised alongside the right to punish and sometimes exercised when it would be wrong to punish. The right to exact compensation, however, need not be regarded as a separate enforcement right. If we say that alongside their ordinary rights persons have more complex rights not to be harmed without compensation being paid them afterwards, we can say that exacting compensation prevents the violation of these rights in exactly the same way that self-defense prevents the violation of simpler rights. Why compensation sometimes can and sometimes cannot be accompanied by punishment—and what distinguishes the two cases—is an extremely difficult question, which lies outside the scope of this paper. For a discussion see Nozick, *Anarchy, State and Utopia*, pp. 57–73.

⁶ A libertarian rights theory of this kind is presented in Immanuel Kant, *The Metaphysical Elements of Justice: Part I of the Metaphysics of Morals*, trans. John Ladd (Indianapolis: Liberal Arts Press, 1965), and discussed by Hart in “Are There Any Natural Rights?”

proper part, and so on). But this is all we need to show why a libertarian rights theory has to contain enforcement rights. Let us imagine that we have discovered that L is the most extensive liberty not containing the liberty to do any enforcing such that every person can have an equal right to exercise all the liberties in L , and no one person's right conflicts with that of any other. Then in deciding whether to grant enforcement rights we are deciding whether to add to L the liberty of removing from other persons the liberty of removing liberties in L from other persons. We have every reason to do this and no reason not to. If we add this liberty—and it is once again best described as the liberty to remove from other persons the liberty of removing liberties in L from other persons—we will be creating a new liberty L' , which contains L as a proper part, and is therefore more extensive than it. But at the same time we will not be subtracting liberties from any other person's liberty L . Although the liberty we are adding conflicts with some liberties of other persons, these are all liberties which have already been excluded from L , and have therefore already been excluded from the protection of their natural right to liberty. Allowing enforcement rights enables us to extend the scope of everyone's right to liberty—which is just what a libertarian rights theory requires us to do—without detracting in any way from the right to liberty of others. And if this is the case, then a libertarian rights theory can give exactly the same justification for these rights as for any other rights it grants.

This justification of enforcement rights, which I have presented so far in a fairly abstract way, applies most directly to the right of self- (or other-) defense. If we give persons the right to use force to prevent rights violations then we are obviously extending the scope of their right to liberty without limiting the right to liberty of anyone else, for no one has the right to violate rights. But it also applies to the right to make and carry out threats which lies behind the right to punish. When we threaten a person with harms if he successfully violates rights we do not remove from him the liberty of violating rights as such. But we do remove from him the more complex liberty of violating rights and not having those harms inflicted on him afterwards. If he does not have the right to exercise the simple liberty he does not have the right to exercise the more complex one either, and in giving other persons the right to remove the more complex liberty from him a libertarian rights theory is once again extending the scope of their right to liberty without in any way detracting from his.

Because it derives the right to punish from a right to make certain threats, the rights-based justification has two attractive consequences which also follow from the retributive justification. The first is that it is never permissible to punish persons who have not violated, or who have not been found by reliable proceedings to have violated, the rights of other persons. Guilt, in other words, is a necessary condition of the permissibility of punishment on the rights-based view. The reasoning leading up to this consequence should be fairly evident. The right to use coercion to prevent others from violating rights only entitles us to make a very specific threat, namely the threat to inflict certain harms on them if they actually

succeed in violating rights, and we could not claim to be carrying out this threat if we inflicted harms on someone whom we did not have reliable reasons to think had violated rights. This first consequence also follows from the retributive justification, but it is a well-known objection to the deterrence justification that no such consequence follows from it. Critics of the deterrence justification often point out that it could license the framing and “punishment” of an innocent man if this would be sufficiently effective in deterring future crimes. The rights-based justification is not open to this objection, for it holds, along with the retributive justification, that guilt is always a necessary condition of the permissibility of punishment. The second consequence is that it is never permissible to punish persons for rights violations unless our intention to punish persons for those violations has been publicly announced in the past. The reasoning leading up to this consequence should also be evident. If punishment is only permissible because it is the carrying out of a permissible threat, then it is only permissible when that threat has actually been made. Punishments for the violation of secret laws, or for the violation of retroactive laws, are never permissible on the rights-based view, though we can easily imagine circumstances in which they would be permissible and even required on the deterrence view, and perhaps even on some retributive views as well.

The rights-based justification, then, has some attractive consequences in common with the retributive justification for the question when punishment is permissible. But when it turns to the question how much punishment is permissible, or how severe a punishment is permissible, it has some consequences in common with the retributive justification and some in common with the deterrence justification. The important thing to realize here is that the enforcement rights that persons have in the state of nature are not unqualified. They are subject to at least two qualifications, and these qualifications place limits on the severity of the punishments which they may inflict in the state of nature, and which their governments may inflict in civil society. To set out these qualifications I will begin by examining some particular cases involving self-defense where I think their intuitive attractiveness is especially evident, and then give them a theoretical justification. I will conclude by showing what the implications of these qualifications are for questions about the morality of punishment, and in particular for questions about the morality of capital punishment.

Let us begin by imagining the following case. One person *X* is trying to violate a fairly unimportant right of another person *Y*, say, the right not to be tickled,⁷ and *Y* is considering how to prevent this. *Y* is not nearly as strong as *X*, so he cannot hope to stop *X* just by resisting him physically. Nor will any threat of *Y*'s deter *X*. But *Y* does have in his hands a pistol with which he can kill *X*. If killing *X* is the only way *Y* can prevent *X* from violating his right not to be tickled, is it permissible for *Y* to

⁷ Those who do not believe there is a right not to be tickled as such can imagine that *X* is trying to violate all the (fairly unimportant) rights that he would violate if he tickled *Y* without his consent.

use his pistol? Locke seems to have thought it is permissible, for in section 19 of the *Second Treatise* he says, “a *Thief*, whom I cannot harm but by appeal to the Law, for having stolen all that I am worth, I may kill, when he sets on me to rob me, but of my Horse or Coat.” But Nozick thinks it is not permissible,⁸ and I think most of us would agree with him. We would insist that there is an upper limit on the amount of coercion persons can use to enforce their rights, and that this limit is lower the less important the rights are which they are enforcing. For *Y* to kill *X* just to prevent him from tickling him is for *Y* quite clearly to overstep a limit which is, in the case of a very unimportant right like the right not to be tickled, very low indeed.

Reflection on this case suggests what I will call an *upper limit* qualification on persons’ enforcement rights. The most natural way for a rights theory to express this qualification is as follows. Although *Y*’s right to enforce his right to ϕ entitles him to act in ways which would otherwise involve violating some rights of *X*’s, it does not entitle him to act in ways which would otherwise involve violating any rights of *X*’s which are more important than his own right to ϕ . In the course of enforcing his right to ϕ *Y* can act in ways which would otherwise involve violating *X*’s right to ϕ , or any rights of *X*’s which are less important than his right to ϕ . So if *X* is trying to kill him *Y* can kill *X* in self-defense, or assault him or tie him up. But he cannot act in ways which would otherwise involve violating any rights of *X*’s which are more important than his right to ϕ . For an important right like the right to life this will not be much of a restriction but for other less important rights it will be. For a very unimportant right like the right not to be tickled, for instance, the upper limit qualification will rule out anything more than the very smallest amount of coercion to enforce it.⁹

Now let us imagine another case. *X* is attacking *Y* with the intention of killing him, and *Y* is considering how to prevent this. He has in one hand a pistol, with which he can kill *X*, and in the other hand a tranquilizer gun, with which he can sedate *X* long enough to make his escape but with which he will not do *X* any

⁸ *Anarchy, State, and Utopia*, pp. 62–63.

⁹ In *Anarchy, State, and Utopia*, p. 62, Nozick gives a different version of this upper limit qualification, using the notion of harm, and saying that the upper limit on the harm we can inflict on an attacker is some function $f(H)$ of the harm H which he is threatening to visit on us, where $f(H) > H$, or at least $f(H) \geq H$. If the notion of harm which Nozick is using here is the ordinary utilitarian notion (as it certainly seems to be—see especially the remarks on 58 and 75) then I am not sure that it is the appropriate one to be using at this point in a rights theory, and I am not sure that, if it is used, it will always yield the right results. I will argue below that the right not to be physically assaulted is not less important than the property right which a person has who owns a weapon, and that as a result my version of the upper limit qualification always allows the victim of an assault to destroy his attacker’s weapon if that is the only way he can prevent himself from being physically beaten. Nozick’s version, however, does not always allow this. Whether I can destroy my attacker’s weapon or not depends at least in part on how much he will be harmed by its destruction. If my attacker is not going to do me a very great harm (I already have one broken arm so another will not be much of an added inconvenience), and if he is very fond of his weapon (it is a family heirloom and its destruction will cause him untold misery), then Nozick’s version of the upper limit qualification says I cannot destroy his weapon to prevent him from breaking my arm. This is surely not what a libertarian rights theory ought to say.

permanent damage. The two weapons will be equally effective in repelling X 's attack and Y knows this. Is it permissible for Y to use his pistol and kill X ? Although Y 's killing X would not violate the upper limit qualification I think most of us would agree that it is not permissible. We would insist that there is another limitation on the amount of coercion Y can use to enforce his rights, one which requires him never to use more than the minimum amount of coercion necessary to prevent the violation of his rights. In this case Y 's killing X would involve more than the minimum amount of coercion, for he can also use the tranquilizer gun on X , and killing him is therefore impermissible.

This second case suggests another qualification on persons' enforcement rights, one which I will call a *minimum necessary* qualification, and which it is most natural for a rights theory to express as follows. Although Y 's right to enforce his right to ϕ sometimes entitles him to act in ways which would otherwise involve violating X 's right to ψ , it only does so when it is not possible for Y to prevent the violation of his right to ϕ just as effectively by acting in ways which would otherwise involve violating only rights of X 's which are less important than his right to ψ . (If it is possible for Y to prevent the violation of his right to ϕ by acting in ways which would not otherwise involve violating any of X 's rights, e.g., by running away, this qualification requires him to run away.)¹⁰ The minimum necessary qualification, and the upper limit qualification as well, can be given a somewhat tidier formulation if we make the following terminological stipulation. Let us say that when Y exercises his enforcement rights against X he always *infringes* some rights of X 's, but he does not always *violate* those rights, for not all rights infringements are morally forbidden. Then the upper limit and minimum necessary qualifications can be read as imposing the following two conditions on permissible infringements of the rights of others:

Y 's right to enforce his right to ϕ (or the right to ϕ of a third party) only permits him to infringe X 's right to ψ if

- (i) X 's right to ψ is not more important than Y 's right to ϕ (or the third party's right to ϕ); and
- (ii) it is not possible for Y to prevent the violation of his right to ϕ (or the third party's right to ϕ) just as effectively by infringing only rights of X 's which are less important than his right to ψ .¹¹

In discussing the upper limit and minimum necessary qualifications I have made extensive use of the notion of the *importance* of a natural right, and there

¹⁰ Nozick discusses a (different) version of this minimum necessary qualification in his "Moral Complications and Moral Structures," *Natural Law Forum* 13 (1968): 1–50. See especially the discussion of Principle VII.

¹¹ Analogues of these two qualifications as they apply to self-defense are recognized in the legal systems of Canada, Britain and the United States, though they are not recognized in Germany or the Soviet Union. For a helpful survey see George P. Fletcher, *Rethinking Criminal Law* (Boston: Little, Brown, 1978), pp. 855–75.

will no doubt be questions about exactly what this notion involves. In speaking of the importance of a right I have intended in the first place to speak of something intuitive. We all have, I trust, an intuitive sense that the right to life is more important than the right not to be physically assaulted, which is in turn more important than the right not to be tickled. But the notion can also be given a formal representation in a libertarian rights theory of the kind we have been discussing. If every right is an instance of the right to liberty, then it seems natural to say that one right is more important than another whenever it is a right to a more extensive liberty than the other. And although comparing the extent of some liberties raises obvious difficulties the following should once again be uncontroversial. If one liberty contains another as a proper part, so that exercising the second liberty always involves exercising the first, but exercising the first liberty does not always involve exercising the second, then the first liberty is more extensive than the second. The ranking procedure which these two suggestions yield is perhaps most usefully put as follows: one right is more important than another whenever violating the first right always involves violating the second, but violating the second right does not always involve violating the first (an example: the right to buy property in Canada is more important than the right to buy property in Alberta because preventing a person from buying property in Canada always involves preventing him from buying it in Alberta, but preventing him from buying it in Alberta does not always involve preventing him from buying it in Canada). This ranking procedure does not generate anything like a complete ordering over rights. It only generates a partial ordering, but the ordering is not so partial as to be useless. It has, for instance, some clear results about a number of rights that are important for questions about self-defense. It holds that the right not to be both tied up and beaten is more important than the right simply not to be tied up, that the right not to have both arms broken is more important than the right not to have one's left arm broken, and that the right not to have property valued at \$100 destroyed is more important than the right not to have property valued at \$1 destroyed. It also has some clear results about a number of rights that are important for questions about punishment. It holds that the right not to be imprisoned for ten years is more important than the right not to be imprisoned for five years, and that the right not to be fined \$100 is more important than the right not to be fined \$1. Most importantly for our concerns, however, it has clear results about the right that is most centrally involved in questions about capital punishment, namely the right to life. On a libertarian view the right to life is the right to exercise the liberty of choosing life over death, and imposes on others the duty not to remove that liberty, as they would do if they forcibly chose death for us. But this means that the right to life has to be the most important natural right there is. Choosing life is choosing to exercise all the liberties we do exercise when we are alive, while choosing death is choosing to exercise no further liberties at all. A person who removes the liberty of choosing life from us is therefore removing all our other liberties from us. In violating our right to life he is violating all our other rights as well, for he is leaving us in a

position where we can never exercise those rights again. Although the proper part ranking procedure has clear results in these areas it does not have clear results in certain others. It does not say anything determinate about the relative importance of property rights and rights not to be physically assaulted, for instance, or of property rights and rights not to be imprisoned. These gaps in the ordering it generates weaken but they do not prevent the operation of the upper limit and minimum necessary qualifications. If property rights and rights not to be physically assaulted are unranked with respect to each other then neither is more important than the other, and persons may if necessary use force against others to prevent them from destroying their property, and destroy others' property to prevent them from assaulting them. Far from being an unwanted result this is one which I think we ought to welcome, for our intuitions seem to support the view that in most cases these two forms of self-defense are, if necessary, permissible.¹² (I say "in most cases" because some may deny that we can inflict serious injuries on another just to prevent him from stealing a small amount of money from us, e.g., \$1. A more finely grained ranking of rights than I have produced might capture this intuition, but I do not know how such a ranking could be generated using purely rights concepts. It is worth pointing out, however, that in civil society there exist law enforcement agencies one of whose functions is to restore stolen property to its owners, and that if these agencies can restore our \$1 to us without infringing as many of the thief's rights as we would if we inflicted serious injuries on him, the minimum necessary qualification requires us to refrain from inflicting those injuries. If we consider the theft of \$1 in the state of nature, where no such agencies exist, it does not seem so implausible to me to say that we can inflict injuries to prevent it, though I think it would be implausible to say that we can kill to prevent it.)

Although the upper limit and minimum necessary qualifications are intuitively appealing we will not be fully entitled to accept them until we have provided them with some kind of theoretical justification. We can do this by extending the libertarian justification of enforcement rights that we have already constructed. This justification says that by adding to the liberty L protected by every person's

¹² In "The Paradox of Punishment," *Philosophy and Public Affairs* 9 (1979–80): 42–58, Alan H. Goldman worries that something like the upper limit qualification forbids us to imprison persons for crimes against property, even though imprisoning these persons is necessary if our laws protecting property are to be effective. This worry only arises because Goldman's version of the upper limit qualification is incorrect. "If we ask which rights are forfeited in violating the rights of others," he says, "it is plausible to answer just those rights that one violates (or an equivalent set)," where equivalence "is to be measured in terms of some average or normal preference scale, much like the one used by the utilitarian when comparing and equating utilities and disutilities" (p. 45). Since most people would prefer losing several thousand dollars to spending five years in prison. Goldman concludes that the upper limit qualification forbids us to give five-year prison terms for thefts of several thousand dollars. Goldman's problem here is similar to Nozick's: he is trying to generate a ranking of rights using concepts which belong properly in a utilitarian rather than a rights-based theory. If he used genuine rights concepts he would find that property rights and rights not to be imprisoned are not ranked with respect to each other, and that the upper limit qualification permits imprisonment for crimes against property.

right to liberty the liberty of removing from other persons the liberty of removing liberties in *L* from other persons, we extend the scope of every person's right to liberty without detracting from the right to liberty of anyone else. But things are not quite as simple as this. When we enforce our rights (or those of a third party) against another person we usually remove from him, not only the liberty of removing liberties in *L* from other persons, but also certain other liberties. In fending off his attack we may prevent him from killing us but we also break his arm; in interfering with his robbery attempt we may prevent him from stealing our neighbor's jewelry but we also confiscate his gun, and so on. These are not reasons for denying the existence of enforcement rights—the argument for that remains intact—but they are reasons for placing certain qualifications on them to ensure that their exercise does not do more to interfere with liberty than it needs to, or more to interfere with liberty than it does to protect it. The most obvious such qualification is the minimum necessary qualification. A person whose enforcement rights are subject to this qualification still always has those enforcement rights. He still always has the right to remove from other persons the liberty of removing liberties in *L* from other persons, but he is now required to make sure that, whenever he exercises this right, he removes as few other liberties from these persons as he can. The minimum necessary qualification does not detract in any way from his right to liberty, for it does not take away from him any of the enforcement rights which the libertarian justification says he has. But it does extend the right to liberty of other persons by making certain interferences with their liberty impermissible. The more difficult qualification to justify is the upper limit qualification, for this qualification does take away some of the enforcement rights which the libertarian justification seems to say a person has. Someone whose enforcement rights are subject to the upper limit qualification rights does not really have those rights when the only way of exercising them would involve infringing rights of other persons that are more important than the ones he is trying to protect (as when the only way of preventing another from tickling him is by shooting him dead). But I think the upper limit qualification can still be shown to follow from the basic principle of a libertarian rights theory. If this principle says that every person has a right to the most extensive liberty compatible with a like liberty for others it will hardly want to allow persons who are enforcing rights to remove more liberty from others than they are trying to protect themselves. But this is just what enforcement rights without the upper limit qualification would allow. They would allow a person protecting a very narrow liberty of his own (say, the liberty to choose not to be tickled) to remove much more extensive liberties from another person, and even to remove all his liberties by shooting him dead. Enforcement rights without the upper limit qualification would sometimes cost more in terms of liberty removed than they would gain in terms of liberty protected, and a theory which wants us to have a right to the most extensive liberty possible will therefore surely insist on the qualification.

Having discussed the upper limit and minimum necessary qualifications in a general way, let us now see what their implications are for questions about

punishment. The qualifications place limits on the severity of the punishments which persons are permitted to inflict in the state of nature, and which their governments are permitted to inflict in civil society. It follows from the upper limit qualification that they are never permitted to inflict punishments which infringe rights that are more important than the ones which the offender has violated, and which they are therefore enforcing. And it follows from the minimum necessary qualification that they are never permitted to inflict punishments which infringe rights that are more important than is necessary to prevent further violations of the right which they are enforcing. If two punishments will be equally effective in deterring violations of this right, they have a duty to impose the less severe punishment; and if no punishments will be effective in deterring violations, they have a duty to impose no punishment at all. Something like this last consequence was accepted by Locke, and the fact that it was suggests that while Locke may not have accepted the upper limit qualification, he did accept the minimum necessary qualification on enforcement rights. When Locke says in section 7 of the *Second Treatise* that persons have the right to enforce the Law of Nature he says they have the “right to punish the transgressors of that Law *to such a degree, as may hinder its violation*” (my emphasis). The suggestion implicit in this remark—that punishment is only permitted when it has some independent deterrent effect—is confirmed at several points in the ensuing discussion. In section 8 Locke says that punishment must be such “as calm reason and conscience dictates, what is proportionate to his Transgression, which is so much as may serve for *Reparation* and *Restraint*,” and that persons in the state of nature have the right to “bring such evil on any one, who hath transgressed that Law, as may make him repent the doing of it, and thereby deter him, and by his Example others, from doing the like mischief.” In section 12 he says that “Each Transgression may be *punished* to that *degree*, and with so much *severity* as will suffice to make it an ill bargain to the offender, give him cause to repent, and terrifie others from doing the like.” There is no suggestion in these passages of the traditional retributive view that there is a “fitting” punishment for every crime, fixed by a “retributive calculus” and independent of any facts about the probable consequences of the punishment. In Locke’s view the permissible severity of a punishment depends entirely on its probable consequences: it is the minimum severity necessary for the effective deterrence of the rights violation for which the offender is being punished.¹³

Of these two consequences, the one that follows from the upper limit qualification also follows from many versions of the retributive justification. Many retributive theorists also hold that there is an upper limit on the severity of the punishments

¹³ Nozick does not, curiously enough, appeal to his version of the minimum necessary qualification when he discusses punishment in *Anarchy, State, and Utopia*, pp. 59–63, but instead presents a version of the “retributive calculus.” Not only does this mark a (to my mind unattractive) departure from Locke, it also weakens the parallel which Nozick draws later on in his discussion between issues about punishment and issues about self-defense.

we can inflict for certain crimes, and that we do wrong if we exceed this limit. But no such consequence follows from the deterrence justification. The deterrence justification permits and even requires as severe a punishment as will best promote the overall good of society, and this punishment can sometimes be very severe indeed. It might well be the case that capital punishment would be an effective deterrent to the crime of shoplifting, and that the benefits to society as a whole of the huge reduction in shoplifting resulting from its imposition would far outweigh the harms to the one or two individuals foolish enough to be caught and executed for shoplifting. Most of us do not think, however, that it could ever be permissible to impose capital punishment for the crime of shoplifting, and take it to be a serious objection to the deterrence justification that it could sometimes require it. The consequence which follows from the minimum necessary qualification also follows from the deterrence justification, but it does not follow from the retributive justification. The retributive justification can require us to impose severe punishments when no further rights violations will be prevented by them, and indeed when no further social good will result from them at all. Some retributive theorists like Kant have of course reveled in this fact, but I think most of us find it repugnant. We think that punishment is only permissible when it does something to promote social good, and take it to be a serious objection to the retributive justification that it requires it even when it does nothing to promote social good.

What are the consequences of the rights-based justification for the special case of capital punishment? Capital punishment infringes the right to life of a criminal, and the right to life is the most important right there is. This means that, given the upper limit qualification, the rights-based justification will only allow capital punishment to be imposed on persons who have violated the right to life of another, that is, it will only allow capital punishment to be imposed for the crime of murder. At the same time, however, given the minimum necessary qualification, the rights-based justification will only allow capital punishment to be imposed for the crime of murder if there is no other less severe punishment that is equally effective at deterring murder. Extensive criminological studies have failed to produce any evidence that capital punishment is a more effective deterrent to murder than life imprisonment, and the rights-based justification will therefore hold that, until such evidence is produced, the imposition of capital punishment for any crime at all is impermissible.¹⁴ This is in my view an attractive consequence, and it is one that also follows from the deterrence justification. But it is not bought at the cost of the many unattractive consequences of the deterrence justification. Many of us believe that if capital punishment is not an effective deterrent to murder then it ought not to be imposed. But we would not want this view to commit us to the simple deterrence justification, with all the unattractive consequences which that

¹⁴ As is often pointed out, the studies have not produced evidence that capital punishment is *not* a deterrent to murder either. But the onus of proof in this question is surely on the defenders of capital punishment to show that it is.

justification has. We would not want it to commit us to the view that capital punishment could be a permissible or even a required punishment for shoplifting, and we would not want it to commit us to the view that it could be permissible or even required to frame and “punish” an innocent man. The rights-based justification allows us to give some weight to the question of deterrence in assessing the morality of capital punishment, without giving it the overwhelming weight that it has in the deterrence justification.

Perhaps the distinctive consequences of the rights-based justification for the morality of capital punishment can best be summarized as follows. Assuming that a retributive calculus will find capital punishment a “fitting” punishment for the crime of murder, the retributive justification holds that it is a necessary and sufficient condition for the permissibility (and even requiredness) of imposing capital punishment on a person that he be guilty of murder. The rights-based justification agrees that this is a necessary condition but denies that it is sufficient; for a punishment to be permissible, it maintains, it must have some independent deterrent effect. The deterrence justification, by contrast, holds that it is a necessary and sufficient condition for the permissibility (and even requiredness) of imposing capital punishment on a person that this punishment have some independent deterrent effect. The rights-based justification once again agrees that this is a necessary condition but denies that it is sufficient; for a punishment to be permissible the person who undergoes it must be guilty of a crime, and guilty of a crime which violated rights at least as important as those which his punishment will infringe. In the rights-based justification conditions which are individually both necessary and sufficient in the deterrence and retributive justifications are made individually necessary but only jointly sufficient, and for this reason the rights-based justification can be said to combine the attractive features of the other two justifications while avoiding their unattractive features. The view that the conditions focused on by the deterrence and retributive justifications are individually necessary but only jointly sufficient for the permissibility of punishment has of course been defended by a number of philosophers. But I think it is only in the context of a rights-based moral theory that this view can be given a theoretical justification, and the attractive features of the deterrence and retributive justifications combined in a manner that is principled rather than ad hoc.

Postscript

One of this chapter’s key claims, that what is most fundamentally justified is not punishment itself but the threat to punish, was defended soon after its publication by Warren Quinn in his essay “The Right to Threaten and the Right to Punish,”¹⁵

¹⁵ *Philosophy and Public Affairs* 14 (1985): 327–73.

which generated considerable discussion. What the original essay called the “upper limit” condition on self-defense I would now call, more standardly, the proportionality condition. But there is a problem in the chapter’s attempt to extend this condition from self-defense to punishment. To yield intuitively acceptable conclusions about punishment, the proportionality condition must be very strict; it must forbid infringements of any rights that are more important than those whose violation is being punished, since only then can capital punishment be restricted to crimes of homicide, as I think most will want it to be. But that strict a condition is not plausible for self-defense, where on most views killing can be permissible to prevent rape, serious bodily injury, or confinement against one’s will for a long period of time; here a defender may infringe a right of his attacker that is somewhat more important than the one he is defending. While both self-defense and punishment are governed by proportionality constraints, the one in the first case seems less strict than what is called for in the second.

Two Kinds of Satisficing

Consequentialist moralities share a general moral structure. They all identify some state or states of affairs as intrinsically good, and characterize the right act in terms of the quantity of good it produces.

The most familiar such moralities are *maximizing* moralities. They characterize the right act as that which produces the most good, or has the best consequences. In these moralities an agent's duty is always to produce the most good possible.

Recently, Michael Slote has defended an alternative that he calls *satisficing* consequentialism. Less demanding than maximizing, it requires only that agents produce consequences that are "good enough."¹ Satisficing consequentialism selects a threshold of goodness in outcomes that is reasonable or satisfactory. Agents are morally bound to aim at outcomes that reach this threshold, but they are not bound beyond that. Although they may, if they wish, bring about outcomes that are more than satisfactory, they need not do so; and, if they do not, they are in no way at fault.

Slote assumes a subjective theory of the good, one on which the good consists in pleasure, the satisfaction of desire—something like happiness. Given this theory, satisficing requires agents to bring about satisfactory happiness. Even with this addition, however, Slote's characterization of satisficing is ambiguous. His general idea that consequences need only be "good enough" admits of two interpretations that Slote does not distinguish, and which, when they are distinguished, appear in a very different light. One is indeed, given a subjective theory of the good, an attractive alternative to maximizing; the other is not.²

On the first interpretation, satisficing selects its threshold of satisfactory goodness without reference to the alternatives an agent has. It selects some *absolute*

¹ Michael Slote, *Common-sense Morality and Consequentialism* (London: Routledge & Kegan Paul, 1985), p. 36.

² Though I cannot argue this here, I do not believe that any version of satisficing is plausible given objective or perfectionist values. With these values, the only plausible structure is maximizing.

level of goodness in outcomes as satisfactory, and requires agents to aim at that. When a situation is and will remain below the absolute threshold, an agent's duty is the same as under maximizing: she must do everything to move it towards satisfactory goodness. Once the threshold is reached, however, her duty vanishes. If a state of affairs is already, by absolute standards, reasonably good, she has no duty whatever to improve it.

The second interpretation is *comparative*. It says that an act's outcome is good enough if it is reasonably close to the best outcome the agent could have achieved. On this interpretation, an agent's duty is always but only to bring about some reasonable percentage of the greatest goodness she can, or to make some reasonable percentage of the largest contribution to goodness she can.

These two interpretations, absolute-level and comparative satisficing, coincide in an important range of cases. These are cases where the situation before a person acts is considerably below the absolute threshold, and he has the option of improving it either to the threshold or beyond it. Many of Slote's central illustrations are of this type. One is a common occurrence in fairy stories. The hero, offered a single wish, asks for a pot of gold, for a million dollars, or simply for enough to make his family and himself comfortably well off for the rest of their lives.³ Maximizing says this hero has acted wrongly, for he did not ask for more money when he easily could have. But both versions of satisficing disagree. The hero's condition after his wish is, considered in itself, reasonably good. It is also, while not the best he could have achieved, reasonably close to the best.

In other cases, however, the interpretations diverge. If the state of affairs before a person acts is below the absolute threshold, and cannot be raised to that threshold, absolute-level satisficing on its own tells him to make it as good as possible, or to bring it as close as possible to absolute satisfactoriness. But comparative satisficing is less demanding. No matter how bad the initial situation, it requires agents only to make some reasonable percentage of the largest improvement in it they can. There are also divergences in the opposite direction. If the state of affairs before a person acts is above the absolute threshold, absolute-level satisficing gives him no duty whatever to improve it. But comparative satisficing on its own still does. It still requires him to do something to improve the outcome, no matter how good that outcome is guaranteed to be.

Given these two interpretations, there are three forms that a satisficing consequentialism can take. It can employ just absolute-level satisficing; it can employ just comparative satisficing; or it can employ both together. In the latter case it will relieve agents twice from the demands of maximizing: when the situation is already above the absolute threshold, it will require nothing of them; and, even when it is below that threshold, it will require only that they do something less than the most possible to improve it.

³ Slote, *Common-sense Morality and Consequentialism*, p. 43.

As I have said, Slote does not notice the difference between the two interpretations of satisficing. Nevertheless, one can determine which form of satisficing consequentialism he accepts. At different points in his discussion he endorses both versions of satisficing, and therefore must accept the third, or doubly satisficing, view.

This is evident first in Slote's general characterizations of satisficing, which alternate between talk of absolute levels of goodness⁴ and talk of "some sort of percentage or other mathematical function of the best results attainable by the agent."⁵ It is also evident in his discussions of examples. One of the first of these runs as follows. "It is mid-afternoon; you had a good lunch, and you are not now hungry; neither, on the other hand, are you sated. You would enjoy a candy bar or Coca-Cola, if you had one, and there is, in fact, right next to your desk, a refrigerator stocked with such snacks, and provided gratis by the company for which you work."⁶ Where maximizing says you are required to get the candy and Coca-Cola, Slote thinks you need do nothing. But this is an absolute-level, not a comparative, claim. On its own, comparative satisficing might say you are required to get one of the candy and Coca-Cola, thus making some percentage of the largest contribution you can to your happiness. If Slote here affirms absolute-level satisficing when it diverges from comparative satisficing, he also does the opposite. "A medic attending the wounded on the battlefield may attend to the first (sufficiently) badly wounded person he sees without considering whether there may be someone in even worse shape nearby."⁷ Similarly, a doctor who wants to help humanity need not go to the country where the suffering is greatest. If he has a particular interest in some other country he may go there instead,⁸ thus making less than the greatest improvement in what will remain, by absolute standards, an unsatisfactory state of affairs.

Is Slote's doubly satisficing consequentialism the most plausible such consequentialism? I do not believe it is. I believe that, while absolute-level satisficing is attractive given a subjective theory of the good, comparative satisficing is not. The most plausible satisficing consequentialism is therefore the first, which employs only absolute-level satisficing.

The problem with comparative satisficing comes in those cases where it makes its distinctive claims. For comparative satisficing holds that, when others are suffering very badly, and nothing we do can relieve all their suffering, we are nonetheless permitted to do less than the most possible to help them. And this is hard to accept intuitively. There are moralities that recognize no duty to relieve others' suffering, and perhaps these moralities are understandable. But, if we admit some such duty, it is hard not to make it a duty to relieve as much suffering as possible. Both absolute-level and comparative satisficing offer some relaxation from the

⁴ Slote, *Common-sense Morality and Consequentialism*, pp. 35, 37.

⁵ Slote, *Common-sense Morality and Consequentialism*, p. 52.

⁶ Slote, *Common-sense Morality and Consequentialism*, p. 39.

⁷ Slote, *Common-sense Morality and Consequentialism*, p. 48.

⁸ Slote, *Common-sense Morality and Consequentialism*, pp. 51–52.

demands of maximizing, but they do so in very different contexts. Absolute-level satisficing offers its relaxation when the global situation is already reasonably good. Comparative satisficing can offer it when the situation is desperately bad; this seems a dubious time for relaxation.

Slote recognizes this feature of comparative satisficing and presents examples to make it plausible: those of the medic on the battlefield and the doctor helping humanity. But both examples are flawed, in ways Slote himself should recognize.

Slote is normally very clear about his aims in defending satisficing. He wants it to be a fundamental normative principle, not a derivative one, by consciously following which we may better satisfy some other (e.g., maximizing) principle. But our response to the medic example turns on treating satisficing as derivative. The reason we think the medic may treat the first sufficiently badly wounded person he sees without considering whether someone else needs him more is strategic. We fear that, by wasting too much time ranking injuries, he will do less good than if he starts treatment immediately. For a true test of comparative satisficing we must imagine that the medic comes to the battlefield, sees right off that one of two soldiers is more seriously wounded, yet attends to the other. If we believe there is some duty to relieve suffering, we will find this behavior hard to justify.

The example of the doctor is similarly flawed. Slote is careful to distinguish satisficing from another device for relaxing the demands of maximizing. Some philosophers propose adding to consequentialism an agent-relative permission, one that lets agents give some extra weight to their own good when it conflicts with others'. As Slote points out, satisficing differs from this agent-relative permission, for it applies even when helping others involves no sacrifice. Consequently, he usually constructs his other-regarding examples so that doing more for others involves no extra cost for the agent.⁹ But the doctor example does not have this feature. If the doctor wants to go to one country, his going to another where the suffering is greater will involve a sacrifice on his part. For a pure illustration of satisficing we must imagine that the doctor goes to a country where he will do less good even though this is not significantly better for him. Again, it is hard to find this acceptable.

Let me summarize. Slote presents satisficing as a unitary idea, but it admits of two interpretations. On the first, an act's consequences are "good enough" if they reach some absolute level of satisfactory goodness; on the second, if they are reasonably close to the best the agent can achieve. Given this difference, there are three possible versions of satisficing consequentialism, two employing just one interpretation, a third employing them both. The most plausible such consequentialism employs just absolute-level satisficing. Comparative satisficing, once distinguished from absolute-level satisficing, has intuitively objectionable consequences.¹⁰

⁹ See, for example, Slote, *Common-sense Morality and Consequentialism*, p. 45.

¹⁰ For helpful comments on an earlier draft, my thanks to Michael Slote.

The Justification of National Partiality

The moral issues about nationalism arise from the character of nationalism as a form of partiality. Nationalists care more about their own nation and its members than about other nations and their members; in that way nationalists are partial to their own national group. The question, then, is whether this national partiality is morally justified or, on the contrary, whether everyone ought to care impartially about all members of all nations. As Jeff McMahan emphasizes in his essay “The Limits of National Partiality,” a philosophical examination of this question must consider the specific features of nationalism as one form of partiality among others.¹ Some partiality—for example, toward one’s spouse and children—seems morally acceptable and even a duty. According to commonsense moral thinking, one not only may but also should care more about one’s family members than about strangers. But other instances of partiality, most notably racial partiality, are in most circumstances widely condemned. Is national partiality more like familial partiality or more like racial partiality? To answer this question, we must know what in general justifies attitudes of partiality. Caring more about certain people is appropriate when one stands in certain special relations to those people. But what are these relations, and to what degree do they hold among members of the same nation? Assuming they are present within families and not within races, to what degree are they present within nations?

In addressing these questions, I will consider only “universalist” nationalism, the view that *all* people ought to be partial to their own nation and conationals. This is a

An earlier version of this essay was presented as a commentary on Jeff McMahan’s “The Limits of National Partiality” at the Conference on the Ethics of Nationalism, University of Illinois at Urbana-Champaign, April 1994. Revised versions of both essays were then published in the volume deriving from the conference. Many of my ideas were stimulated by McMahan’s fine essay; I am also grateful to him and to Robert McKim for helpful comments.

¹ Jeff McMahan, “The Limits of National Partiality,” in *The Morality of Nationalism*, ed. Robert McKim and Jeff McMahan (New York: Oxford University Press, 1997): 107–38.

more interesting and plausible position than the “particularist”—one could equally well say “chauvinist”—view that only one’s own nation, say, only Canada, deserves special loyalty. And I will consider only intrinsic justifications of nationalism. There are various instrumental arguments for national partiality, ones claiming that, starting from impartial moral principles, we can show how people’s being partial to their conationals will have good effects impartially considered. I do not find these instrumental arguments very persuasive. In any case, the more interesting philosophical question is whether national partiality can be justified noninstrumentally, or at the foundational level of morality. Many people believe that familial partiality is justified not just as a means to benefits for all but intrinsically or in itself. My question will be whether national partiality can be justified in the same foundational way.

My discussion will cover three separate topics. First, I will challenge one widely accepted view about the moral foundations of nationalism. Second, I will suggest that a full discussion of nationalism must recognize that it has two components, which raise distinct moral issues. Finally, I will sketch a moral defense of one of these aspects of nationalism. This defense will concede that along one important dimension the relations among conationals have less of the character that justifies partiality than do the relations among family members, but it will argue that along another dimension they have roughly as much. The result is not that we should be as partial to our conationals as we are to our children—that would be absurd—but that we may properly be partial to some degree.

Nationalism and “Embedded Selves”

I have said that the moral issues about nationalism turn on whether certain relations hold within national groups. But many writers connect these issues to more abstract debates about the nature of morality and of moral agents. They say the impartialist view that we ought to care equally about all humans goes with the “Enlightenment” conception of morality as universal and impartial, whereas the defense of nationalism goes with a different “particularist” or “communitarian” conception of morality. According to this latter conception, moralities necessarily arise within the life of particular communities and therefore inevitably distinguish centrally between members and outsiders, requiring a certain priority for the interests of the former. David Miller expresses this kind of view. He says that moral impartialism sees the subject “as an abstract individual, possessed of the general powers and capacities of human beings—especially the power of reason,” whereas a nationalist ethic sees the subject as “partly defined by its relationships, and the various rights, obligations, and so forth that go along with these, so these commitments themselves form a basic element of personality.”² But although

² David Miller, “The Ethical Significance of Nationality,” *Ethics* 98 (1988): 649–50. The same view is defended in his *On Nationality* (Oxford: Clarendon Press, 1995).

nationalism is often said to rest on these communitarian ideas about morality—let me summarize them in the slogan that “moral selves are embedded”—I do not see any connection between the two. Despite their prominence in the recent literature on nationalism, claims about the “embedded self” are not relevant to the intrinsic justification of national partiality. I can detect two arguments in defense of nationalism that may be suggested by the talk of embedded selves, which I will call the *cultural perfectionist* argument and the *metaethical particularist* argument. The cultural perfectionist argument does not go far enough to justify a universalist ethic of nationalism; the metaethical particularist argument goes too far. Let me start with the cultural perfectionist argument.

Perfectionism as a general normative view holds that the good for human beings consists in developing their “nature” or “identity.”³ More specifically, it holds that certain properties make an individual what he or she is and thereby constitute that person’s nature or identity, and that his or her good consists in developing these properties to a high degree. In many classical versions of perfectionism—for example, those of Aristotle, Marx, and Nietzsche—the relevant nature is generic human nature, one shared by all human beings. But those who talk of “embedded selves” sometimes suggest a different perfectionist view. According to this view, human beings have natures or identities based on their membership in particular cultures, and their good consists at least partly in developing these narrower cultural identities. According to this “cultural” perfectionism, I as a Canadian have a specifically Canadian identity, a German has a German identity, and in each case our good consists at least partly in developing this cultural identity. One argument suggested by the talk of “embedded selves” is that this cultural perfectionism provides the justification for national partiality. If human beings had just an abstract or common human nature, this argument runs, a purely impartialist or cosmopolitan morality would be reasonable. But if, instead, their identities depend on their belonging to particular cultures, morality demands that they be specially loyal to those cultures.⁴

Though this is a less central point, I do not believe that cultural perfectionism is a very plausible version of perfectionism. In the most attractive versions of this general normative view, the properties that it is good for a human to develop constitute his or her identity in a strict or metaphysical sense. They are essential to the person in the strong sense that he or she could not exist as numerically the same individual without having these properties. This condition is satisfied by the classical perfectionisms of Aristotle, Marx, and Nietzsche; since I am essentially a human, no being that did not have the properties that make humans human could be identical to me. But the condition is not satisfied by cultural perfectionism.

³ See my *Perfectionism* (New York: Oxford University Press, 1993).

⁴ This argument is suggested in Michael Sandel, “The Procedural Republic and the Unencumbered Self,” *Political Theory* 12 (1984); and Yael Tamir, *Liberal Nationalism* (Princeton: Princeton University Press, 1993), ch. 1.

I was born in Canada and raised in Canadian culture. But we can easily imagine a different course of events, one in which, a few months after my birth, my parents return to their native Czechoslovakia and raise me there. And what we imagine in this course of events is that *I*, the very same individual, am raised in Czech culture. My being a Canadian, therefore, is not metaphysically essential to me and constitutes my identity in only a weaker, nonmetaphysical sense. And nonmetaphysical identities, it seems to me, cannot generate a plausible version of perfectionism.

As I said, however, this is a less central point. What is more important is that even if we accept cultural perfectionism, it does nothing to justify national partiality. Let us grant that humans have different goods based on their membership in different cultures. How does it follow that I should care more about the achievement by my conationals of their specific cultural good than about the achievement by people in other cultures of their specific good? What rules out the view that I should care impartially about all people's realizing their different cultural identities—that is, what rules out an impartialist cultural perfectionism? Such a perfectionism would recognize that the good of Canadians is different from the good of Germans but direct both Canadians and Germans to care equally about both. How does accepting cultural perfectionism as a general position rule this specific view out?⁵ I am not suggesting that the writers who embrace cultural perfectionism do so in an impartialist way. Most, it seems to me, endorse national partiality. But the partiality they affirm does not follow from their cultural perfectionism, which is equally consistent with an impartialist approach. Their position therefore combines cultural perfectionism with claims about partiality that are independent of any ideas about cultural identities and cannot be justified by them.

Cultural perfectionists may object that I have ignored a crucial feature of their argument. This argument does not claim only that people in different cultures have different identities; it claims, beyond that, that those identities involve, as one component, a demand for partiality toward the culture's members. Thus my identity as a Canadian demands partiality toward Canadians, a German's identity demands partiality toward Germans, and neither of us can fully achieve his or her good by following an impartial morality.

If it takes this form, the cultural perfectionist argument requires a strong additional assumption. To show that national partiality is justified in *every* culture, it must assume that *every* culture involves as one component a demand for partiality,

⁵ As an analogy, consider a different version of perfectionism defended by Rousseau, Humboldt, and Mill. They hold that the nature whose realization constitutes a person's good is not one shared with all other humans or even one shared with all members of one's culture but rather a nature distinctive of that person as an individual. Each person has a unique individual identity, by realizing which the person achieves "individuality." But do these writers say that each person ought to care only or even more about his or her own achievement of individuality than about other people's? They do not. They say that each person ought to care impartially about the achievement of individuality by all and to support those institutions, especially liberal institutions, that will permit individuality for all. But if this individualist perfectionism is compatible with full impartiality, surely cultural perfectionism is as well.

so there could never be a culture of pure impartialists. I find this assumption dubious, but let us grant it and ask what follows. If my identity involves as one component a demand for national partiality, I cannot fully achieve my good if I do what is right by impartialist standards. But this is no embarrassment or difficulty for an impartialist cultural perfectionism. It is merely one instance of the familiar fact that to do what is right, or has the best consequences impartially considered, agents must sometimes sacrifice some of their own good. Doing what is right often involves omitting what is best for oneself; here it involves omitting that part of one's good that consists in being partial. This familiar fact does not tell against impartial cultural perfectionism, and there is still no justification for national partiality.⁶

The difficulty with the cultural perfectionist argument is that it operates at the wrong level to justify national partiality. The affirmation of partiality concerns the *form* of an ethically appropriate concern. It says that whatever people's good consists in, we should care more about our conationals' good than about other people's. But cultural perfectionism makes claims about the *content* of our ethical concern, or about what people's good consists in. And no claims about what people's good consists in can justify the idea that we ought to care more about some people's good than about others'.

The second argument suggested by the talk of "embedded selves," which I called the metaethical particularist argument, does address issues about form. It claims that an impartialist morality, one requiring all humans to care equally about all others, is inconsistent with the true nature of moral codes and principles. These codes and principles, the argument says, always arise within particular cultures; they are addressed to the members of a culture as having the particular cultural identities they have and as occupying particular roles within that culture. Morality is always *our* morality, in these circumstances here. This means that the standpoint presupposed by impartial morality—outside all cultures and making judgments about them all—is not available. Morality must be partial because the impartialist alternative is conceptually incoherent.

The problem here is that the particularist argument excludes not only impartialist morality but also a universalist ethic of nationalism. For universalists, too, make claims about what is right in all cultures, namely, partiality toward them; their judgments, too, do not arise from their particular culture but apply equally to all cultures. Consider Alasdair MacIntyre's lecture, "Is Patriotism a Virtue?" As its title indicates, this lecture asks a question about the value of patriotism in all cultures everywhere. And MacIntyre ties an affirmative answer to this question to

⁶ Could the cultural perfectionist claim that the demand for partiality is not just one component of every cultural identity but an essential component, so that if I do not care more about my conationals I do not realize my cultural identity to any degree at all? This claim surely is, as a claim about identities, utterly implausible. And even if it is accepted, it still does not show why, at a foundational level, I should care more about the realization to some rather than no degree of my nation's identity than of others'.

what looks like a particularist metaethical view. On the view that underwrites patriotism, he writes, we never learn “morality as such, but always the highly specific morality of some highly specific social order.” Later he claims that impartialist morality requires something that cannot be justified, namely, that we “assume an abstract and artificial—perhaps even an impossible—stance, that of a rational being as such, responding to the requirements of morality not qua peasant or farmer or quarterback, but qua rational agent who has abstracted him or herself from all social particularity.”⁷ But someone who really accepted this metaethical particularism could not ask the universalized question of MacIntyre’s title. Such a person could only write a lecture titled “Is Canadian Patriotism a Virtue in Canada?” (if he or she was Canadian) or “Is German Patriotism a Virtue in Germany?” (if he or she was German). This person would not make any claims but would deny the intelligibility of claims about the value of patriotism in cultures other than his or her own. Any such universalist claims, no less than those of impartialist morality, issue from a standpoint that the particularist says is not available—namely, one abstracted from any particular social identity and addressed to all humans or all members of cultures as such. As I have said, the interesting affirmation of national partiality is the universalist one; it is also the one that all writers on this subject discuss. But this universalist affirmation cannot be supported by metaethical particularism; on the contrary, it is excluded by it.

Let me summarize my discussion of the “embedded self” by introducing some technical terminology. The interesting nationalist doctrine is both universalist and agent-relative. It is universalist because it claims that partiality toward one’s nation and conationals is appropriate for all humans in all cultures. It is agent-relative because it says that what different humans should be partial to is different—namely, their *own* conationals. The cultural perfectionist argument does not go far enough to justify this doctrine, because it does not rule out an impartialist view according to which we ought to care equally about the realization of their different cultural identities by people in all the cultures in the world. The metaethical particularist argument goes too far to justify the doctrine, because it rules out not only impartialist but also all universal moral claims. Neither of the two arguments that I can see suggested by the talk of “embedded selves” does anything to justify national partiality. The real issues about the ethics of nationalism do not concern the nature of morality or of the self but are those I introduced at the start of this chapter. Assuming that special relations sometimes justify partiality, are the relations among conationals of the kind that do justify partiality, like those among family members, or of the kind that do not, like those among members of a race?

⁷ Alasdair MacIntyre, *Is Patriotism a Virtue?* Lindley Lecture (Lawrence: University of Kansas, 1984), pp. 9, 12. A weaker version of metaethical particularism is defended by Michael Walzer in *Spheres of Justice* (New York: Basic Books, 1983), chs. 1–2, and *Interpretation and Social Criticism* (Cambridge, Mass.: Harvard University Press, 1987).

Nationalism and Impersonal Goods

My second topic is the content of national partiality, or exactly what nationalists are partial to. Many writers speak simply of being partial to one's nation without explaining further what that means. Some speak, more specifically, of being partial toward one's conationals—that is, of giving more weight to the interests of individuals in one's nation than to those of other individuals. This is certainly one aspect of nationalism, but I believe there is often another aspect.

In a number of writings Charles Taylor has emphasized the importance of cultural survival as a good and value for minority groups. In *Multiculturalism and "The Politics of Recognition,"* for example, he writes: "It is axiomatic for Quebec governments that the survival and flourishing of French culture in Quebec is a good."⁸ Noting the importance of this insight, McMahan says it shows how for participants in a culture its survival has "impersonal value."⁹ I agree that in one important sense the survival of a culture is an "impersonal" value or good, but in another sense, which seems to be the one McMahan has in mind, it is not, or is not most importantly, impersonal.

The survival of a culture is an impersonal good in the sense that is not reducible to the goods of individual persons, or to goods located in individual persons' lives. Consider francophone Quebecers who care deeply that there be a French culture in Quebec three generations from now. Do they believe that the survival of French culture is a good because better human lives will be lived if French culture survives than if it does not? Do they believe, more specifically, that their great-grandchildren will lead better lives if they are born and raised in a French culture than if, that culture having disappeared, those great-grandchildren are born and raised as full members of an English culture? I do not believe these Quebecers need or even should, if they wish to avoid chauvinism, believe this. They should grant that after enough time the disappearance of French culture would not be worse for persons in the sense of making the lives lived by persons worse. If, despite this, they continue to view the survival of their culture as a good, they must view it as an impersonal good in the following sense: it would be better if French culture survived even if this would not make the lives persons live more valuable.¹⁰

⁸ Charles Taylor, *Multiculturalism and "The Politics of Recognition,"* ed. Amy Gutmann (Princeton, N.J.: Princeton University Press, 1992), p. 58.

⁹ See Jeff McMahan, "The Limits of National Partiality."

¹⁰ The importance of goods that are impersonal in this sense for the morality of war is brought out in Jeff McMahan and Robert McKim, "The Just War and the Gulf War," *Canadian Journal of Philosophy* 23 (1993): 522. Note that what Derek Parfit calls an "impersonal," as against a "person-affecting," principle of beneficence does not involve reference to impersonal goods in my sense (see his *Reasons and Persons* [Oxford: Clarendon Press, 1984], pp. 386–87). Parfit's "impersonal" view holds that the best outcome is the one in which the best lives are lived, even though the people living those lives may be numerically different from the people in alternative outcomes. (Numerical nonidentity is especially likely when the outcomes involve large-scale and long-lasting changes, as the survival and disappearance

Valuing cultural survival in this way does not require the metaphysical view that cultures or nations exist separately from, or over and above, their individual members. It is fully compatible with the reductionist view that facts about nations consist entirely in facts about individuals and the relations between them. According to this reductionist view, for French culture to survive in Quebec is only and entirely for individuals in Quebec to live and interact in certain ways. But while holding that the *existence* of a culture is reducible to facts about individuals, a nationalist can deny that the *good* of the culture's existing is reducible to the *goods* of individuals. The fact that people interact in certain ways can have a value that is separate from the values present in their individual lives.¹¹

Cultural survival, then, is an impersonal good in the sense that it does not consist in the goods of individual persons. But the word "impersonal" is often used in another sense, one equivalent to "impartial." In this sense, an impersonal good is one it is appropriate for all agents to desire and pursue and to weigh impartially against other similar goods. This seems to be the sense McMahan has in mind when he calls cultural survival an "impersonal value." He introduces the topic of survival while discussing the instrumental arguments that can be given, from an impartialist standpoint, for endorsing some degree of national partiality, and he considers it alongside a value that cannot but be impersonal in this second sense—namely, that of the overall cultural diversity of the world. But it seems to me that cultural survival is valued by nationalists, and is thought by them appropriately valued, in a highly partial way. Who is it who cares about the survival of French culture in Quebec? It is surely, above all, francophone Quebecers. And they do not care about their culture's survival only in an impartial way, or merely as contributing to a universal good such as overall cultural diversity. If they did, they would gladly accept the disappearance of French culture in Quebec if that somehow allowed the survival of two other cultures elsewhere in the world. This is not their attitude; they care specially about the survival of *their own* culture. In the same way, it seems to me, people outside a culture do not have nearly as much reason to care about its survival as a good. McMahan writes that people outside a culture "are capable of appreciating its intrinsic value" and of "perceiving in a particular alien culture a variety of merits that may not be replicated in any other culture." But these remarks, though true, do not suffice to establish the appropriateness of impartial concern for another culture. I can appreciate that the well-being of someone else's children is a good while believing that I ought to care much more

of a culture do.) But Parfit's "impersonal" view still holds, with other versions of beneficence, that the relevant goods in the different outcomes are all states of individual persons. It is this latter claim that the affirmation of what I call impersonal goods denies. This affirmation rejects not only person-affecting beneficence but also the individualism about value still present in Parfit's "impersonal" view.

¹¹ If the impersonal view accepts metaphysical reductionism, it embodies G. E. Moore's principle of "organic unities," according to which the value of a whole need not equal the sum of the values its parts would have if they existed alone. See Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), ch. 6.

about my own children's well-being. And in my view commonsense nationalism does not give people outside a culture much moral duty to care directly about the culture's survival. This is obscured in many actual situations by the fact that the members of the culture do desire its survival. Thus if francophone Quebecers care deeply about their culture's survival, this gives other people, and especially anglophone Canadians, a reason of a more familiar kind to support measures that will ensure the culture's survival—namely, that Quebecers desire it. But what if a majority of Quebecers ceased to care about their culture's survival and instead preferred assimilating into English culture? In this situation I believe Quebecers in the minority would still feel a strong duty to fight for their culture and to try to persuade the majority to change their minds. But non-Quebeckers would surely not feel any such strong duty, nor would they be failing in not feeling it. They might appropriately feel some mild regret about the loss of a distinctive culture and the loss of some overall diversity in the world, but they would not feel strongly bound to prevent the assimilation—for example, by offering subsidies to Quebecers who retain their French culture. When it is considered in itself and apart from the desires it gives rise to in members, the survival of a culture does not seem to be something that, according to commonsense nationalism, nonmembers have a strong reason to care about or pursue.

I have suggested that cultural survival, though an impersonal good in the sense that it is not reducible to the goods of individuals, is the object of highly partial attitudes. The same can be true of other impersonal goals associated with a culture. For example, nationalists can care that their cultures not only survive but also achieve the full flowering or self-expression that comes through sovereignty and independent statehood. In this case the importance of the impersonal good may be harder to see because there can also be personal goods at stake in sovereignty. Thus nationalists may believe that the individuals in their culture will engage in more valuable political activity or live under more culturally sensitive institutions if their government is entirely their own. But if it is possible to value the survival of one's culture apart from any benefits to individuals, it is surely possible to value sovereignty and statehood in the same way, and I think those active in independence movements do commonly have this impersonal desire. They value their nation's sovereignty, as they value their culture's survival, as something good partly in itself. Thus a central force in the Quebec sovereignty movement has been the desire that francophone Quebecers affirm their status as *un peuple* by establishing their own nation-state. In fact, nationalists can have many impersonal goals that they value in a partial way: that their culture flourish in the arts and sciences, that it be economically vigorous, that it produce athletes who win medals at the Olympics. Beyond this, nationalists can have impersonal political goals that they value partially: that their nation occupy a large territory, that it be militarily powerful, that it dominate its neighbors and even dictate to the world.

In this list of impersonal goals, there is a large moral difference between the innocuous first goal, cultural survival, and the politically threatening ones that come later,

such as military power and world domination. But this is nothing new in the study of nationalism, which is often described as Janus-faced, attractive in some forms and terrifying in others. And our responses to the list may be guided by the view, which many writers on this subject endorse, that any acceptable form of national partiality must be constrained by respect for the basic rights of all individuals, no less in other countries than in one's own. One may pursue one's own nation's good and do so in preference to other nations' good but only in ways that respect fundamental rights.¹² As it happens, the more acceptable impersonal goals, such as cultural survival, can usually be pursued successfully without violating anyone's rights, whereas it is hard to see territorial expansion or world domination achieved without violating rights. The different impersonal goals may differ morally not so much in themselves, therefore, as in the means likely to be necessary for their achievement.

I do not claim that every form of nationalism involves concern for impersonal goods; some nationalists may favor only the interests of their individual conationals. But it seems to me that the two forms of partiality often go together, and I will therefore define *full-blooded nationalism* as combining a greater concern for the impersonal goods of one's own culture, such as its survival and flourishing, with a greater concern for the interests of one's conationals. In a phrase I have used above, full-blooded nationalism involves partiality both toward one's nation, seen as having certain impersonal goods, and toward one's conationals. If this characterization is correct, it has an important implication for the morality of nationalism.

If full-blooded nationalism involves two components, a successful moral justification of it must address both. It must show the appropriateness of partiality toward one's conationals and also toward one's nation's impersonal good. Here the difficulties facing the two justifications seem interestingly different.

Consider, first, partiality toward one's conationals. There is no doubt that one ought morally to care about one's conationals; they are people, and one ought in general to care about people. The difficulty is to show why one should care more about these people than about others who are not members of one's nation, or why partiality toward this particular group is appropriate. In the situation where partiality seems most clearly justified, that of the family, it rests on a special relationship between people that is both rich and intense. The members of a family care deeply about each other, have lived together for many years, and have to a significant degree shaped each other's characters. Their interactions have been as close as people's typically ever are. But the relations among conationals are nothing like this. I have never met the vast majority of my fellow Canadians and do not know who they are; the causal links between our lives are tenuous at best. Especially worrisome is the fact that these links do not seem closer than my links with many non-Canadians—for example, with Americans living just across the Alberta-Montana border. In fact, with respect to closeness, the relations among conationals seem comparable to

¹² See, for example, Jeff McMahan, "The Limits of National Partiality."

those among members of a race, who likewise mostly have not met. If the relations between conationals hold only to a limited degree, and not much more than between non-nationals, how can they justify any substantial degree of partiality?

The justification of the second form of partiality, toward one's nation's impersonal good, faces the opposite difficulty. Here there does not seem to be a large problem about justifying the attitude's partiality. Only one culture or nation in the world is mine; all the others are not mine. This is not just a small difference in degree but a large difference, perhaps a difference in kind. So if the justification of strong partiality requires a large difference in linkage or connectedness, we have that here. The problem, rather, is to show that impersonal goods are morally appropriate objects of any concern in the first place. What can be called "individualist" theories of the good deny this. Individualist theories hold that the only goods there are, and thus the only objects of rational concern, are personal goods, or the goods of individuals.¹³ According to individualism, nationalists who value the survival or flourishing of their culture apart from any effects on individuals are being irrational and fetishistic. Their attitude is objectionable not because of its partiality but because of its object, which is not a genuine good because it is not a feature of individuals' lives. Nor is it only individualism in the strict sense that counts against the second form of partiality. A more moderate view allows that there can be impersonal goods and rational concern for them but insists that these goods are always relatively minor and the concern they call for always of less weight than the concern required for individuals. According to this moderate view, a partial attitude toward one's nation's impersonal good is allowed but not in a strength that often allows promoting that good at the expense of benefits to individuals.

To summarize: If there are two forms of national partiality, they need two justifications, and the difficulties facing these justifications are different. That one should care somehow about one's conationals is not in doubt; the question is whether it is right to care more about them than about non-nationals. As for a nation's impersonal good, if some concern for it is appropriate, it seems plausible that this is a highly partial concern. The difficult question here is whether that initial concern is appropriate: whether impersonal goods are worth caring about or whether the only, or only important, goods are those of individuals.

Partiality and History

Having suggested the importance to nationalism of impersonal goods, I will now set them aside and consider the more commonly recognized aspect of

¹³ Individualism is affirmed in Avishai Margalit and Joseph Raz, "National Self-Determination," *Journal of Philosophy* 87 (1990): 439–61; and Tamir, *Liberal Nationalism*, pp. 83–84. Interestingly, Margalit and Raz allow that the interest of a group is not reducible to the interests of its members (p. 450) but insist that only individual interests are relevant to the justification of national rights.

nationalism: partiality toward one's conationals. This partiality has many more specific manifestations. Nationalists typically care much more about relieving economic hardship within the nation than outside it; compare what nations spend on domestic welfare programs with what they spend on foreign aid. Nationalists also want immigration policy decided primarily by considering the effects on people already within the nation rather than on those who want to join. These various positions may receive some support from concern for impersonal goals like the nation's flourishing as a collective, but they are primarily directed at individuals. Setting aside the impersonal component of nationalism, therefore, I will consider the moral justification of partiality toward one's individual conationals. When partiality toward certain individuals is justified, it is because certain special relations hold between oneself and them. To what degree do these relations hold between members of a nation?

Because the arena in which partiality seems most clearly justified is the family, defenders of nationalism often try to assimilate the relations among conationals to those among family members. As we have seen, however, this assimilation is problematic; especially in the degree of interaction they involve, nations are not like large families. To many writers, therefore, it has seemed that the degree of national partiality that is justified is even in the most favorable circumstances much less than most nationalists desire.

In this section I will sketch a reply to this widespread skepticism about national partiality. This reply concedes that along one important dimension the relations between conationals have much less of the character that justifies partiality than do familial relations, but it claims that along another dimension, which most writers ignore, they have roughly as much.

First, however, I must state a presupposition of my argument: that the basis of partiality among conationals must be an objective rather than a subjective relation and, in particular, cannot be just the fact that conationals care more about each other than about non-nationals. It may be, as is sometimes argued, that certain subjective facts—that is, certain attitudes on the part of individuals—are necessary for a nation to exist. For example, it may be that individuals must view membership in a group as an important part of their identity before the group constitutes a nation. But questions about when a nation exists are different from questions about when its members should be partial toward each other, and the latter questions cannot turn on mere facts about caring. There are two decisive arguments for this conclusion.¹⁴ One is that a purely subjective basis could not rule out the racial partiality that most of us find morally offensive. The fact that racists care more about people with their own skin color would by itself make it right for them to do so. The second argument is that a subjective basis cannot justify what nationalists typically affirm—namely, a duty to favor one's conationals that is binding even on

¹⁴ See Jeff McMahan, "The Limits of National Partiality."

those who do not now care about their conationals. I will assume, then, that the basis of national partiality must be some objective relation—that is, some relation that holds independently of people's attitudes. To determine which relation this is, we must look more closely at the objective side of personal or familial relations.

Consider my relation to my wife. If I love her specially, it is partly for certain qualities that she has. Some of these qualities I am attracted to without judging them to be intrinsically good, such as her appearance and the sound of her voice. Others I do judge to be good, such as her trustworthiness, her intelligence, and her concern for other people. Especially with these latter qualities it is important that my beliefs about them be true—that she, in fact, have the qualities, and that they truly be good. But even if all my relevant beliefs are true, my wife's having these qualities does not explain all my emotional attachment to her. If it did, I would abandon my wife the moment someone else came along with the same properties to a higher degree. Or if, just before dying, my wife had a clone of herself made to stay with me, I would think myself no worse off for the exchange. But of course I would not trade in my wife in this way. Though I love her partly for her qualities, I do not do so in a way that would accept substitution. I also love her, in the common phrase, "as an individual," or for herself.

What does it mean to love a person "as an individual"? In my view, it does not mean loving a person apart from any qualities at all but rather loving the person for qualities that no one else can share. More specifically, it involves loving the person for certain historical qualities, ones deriving from his or her participation with one in a shared history. Thus I love my wife not only as trustworthy, intelligent, and so on but also as the person who nursed me through that illness, with whom I spent that wonderful first summer, and with whom I discovered that hotel on Kootenay Lake. These historical qualities focus my love on my wife as an individual, since no substitute, not even a clone, can be the very person who did those things with me.

A highly romantic view of love and friendship holds that once these historical qualities are established they entirely determine the relationship, which should therefore never end and always imposes duties of partiality. This is the view expressed in Shakespeare's line: "Love is not love/Which alters when it alteration finds." But I think most of us believe that historical qualities, though part of the basis for love and friendship, are, again, not the entire basis. If my wife changes radically, losing the general or shareable qualities I admire and taking on ones I find despicable, I will no longer feel attached to her or bound by duties of partial concern. My love, in other words, has a dual basis. My wife's role in a shared history with me explains why I love her more than other people with similar general qualities, but her general qualities matter, too. If those qualities changed enough, our history would not be a sufficient basis to maintain my love or to continue to demand partiality toward her.

We can see the same dual basis in nationalists' attachment to their nation and conationals. Nationalists are, first, attracted to their culture and the activities that

define it, thinking them to a considerable degree good. They need not believe that their culture is superior in the sense of being the single best in the world. That chauvinist belief would not be credible and, in any case, would justify not universalist nationalism but the belief that everyone in the world should promote the one best culture. Instead, nationalists need only believe that their culture is one of perhaps many in the world that are good. What attaches them specially to this culture and its members are historical facts: that this is the culture *they* grew up in, that their conationals share *with them* a history of being shaped by, participating in, and sustaining this culture. The favorable evaluation of their conationals' cultural activities is a necessary basis for this nationalist attachment, but it is not sufficient. There is also, and distinguishing their conationals from other people whose culture is equally good, the crucial fact of a shared cultural history.

This dual basis can lead to conflicts about national attachment. As Yael Tamir writes, "Citizens of a state involved in an unjust war may be torn between the feeling that they have an associative obligation to serve in the army together with their enlisted fellows, and their commitment to a moral code dictating they should refuse."¹⁵ In the situation Tamir describes, the citizens' state is not now good; it has at least some general qualities that are evil. But the citizens are still historically connected to this state as the one they grew up under. How they resolve this conflict depends on which of the two bases of national attachment they find more important, which in many particular cases will depend on how evil their state currently is. If it is not irredeemably evil, the citizens may continue to feel special duties toward it and work harder to reform it than to reform other equally evil states elsewhere. But if their state degenerates too far, their historical connection to it may be outweighed and their feelings of national attachment, like love for an individual whose character has changed utterly, may end.

If national attachment rests partly on the belief that one's culture is good, it is important that that belief be true, which requires the culture to be, in fact, good. This is one point where evaluative considerations bear on the justification of national partiality, but there is another point as well. Considerations about good and evil also help determine when a shared history is of the right kind to justify partial concern and, when it is, what degree of partiality is justified.

Consider again a personal relationship like that between spouses. Here the shared history is predominantly one of mutual benefit or beneficence; two people have helped each other through difficult times and also shared good times, giving and taking pleasure in each other's company. And I think a history of reciprocal benefit—or, alternatively, one where people have jointly benefited others, such as the students in a school where these people taught—can be a legitimate basis of partiality. The same is true of a history of shared suffering; people who lived in the same barracks in a Nazi labor camp and suffered the same evils there can

¹⁵ Tamir, *Liberal Nationalism*, p. 102.

appropriately feel on the basis of their shared history some greater concern for each other's well-being. But I do not think a shared history justifies partiality when it is a history of doing evil, as for former members of an SS unit that ran and terrorized a labor camp. Many of us find something obscene in the idea of nostalgic reunions, even at this late date, of former SS colleagues, and there is a similar obscenity in the idea of partiality toward former SS colleagues. If an SS veteran receives a letter from one of his former colleagues claiming financial hardship and requesting a loan of \$1,000, should he feel a special duty to honor the request or to help his former colleague before helping others who are equally in need? It seems to me that he should not, even if his former colleague is now morally reformed. If anything, given the evil of the history they share, he should feel a duty not to associate with his former colleague and should contribute first to others who did not participate in that aspect of his past. Whereas a shared history of doing good or suffering evil can justify duties of partiality, a shared history of doing evil cannot.

These points suggest a general account of the basis of duties of partiality. Some activities and states of people, most notably their doing good or suffering evil, call for a positive, caring, or associative response. Others, such as their doing evil, call for a negative or dissociative response. Partiality between people is appropriate when they have shared in the past in the first kind of activity or state. For example, if two people have a shared history of doing good, either reciprocally or to others, partiality between them in the present is a way of honoring that good fact about their past. (This is why partiality among former SS colleagues is troubling; it seems to honor a past that properly calls for dishonor.) One should, in general, care more about people who have shared with one in activities and states that call for a caring response. This account does not claim to justify partiality of concern as a general moral phenomenon; on the contrary, it assumes it. It assumes that one has a special duty to honor past doings of good or sufferings of evil *that involved oneself*. But it does give particular duties of partiality a more abstract basis. In the many realms where partiality is appropriate—the family, private clubs, perhaps the nation—it is an appropriate response to a history that joins oneself and other people in activities or states that are good or that call for association.

This general account can explain our attitudes to racial partiality. As McMahan notes,¹⁶ while we condemn racial discrimination by members of a dominant racial group, we often think it appropriate for minority races to celebrate their distinctness and even to implement discriminatory policies that benefit their members at the expense of others. In current conditions, black and aboriginal solidarity movements have a different moral status than white supremacy movements. The explanation, I would argue, is that minority racial groups have a shared history of the kind that makes partiality morally appropriate—namely, a shared history of

¹⁶ See Jeff McMahan, "The Limits of National Partiality."

suffering evil because of one's racial membership. But the history of dominant racial groups, which is largely one of oppressing the minority, is not of the kind that justifies partiality. Among members of the minority, there is a shared history that morally warrants partiality toward other members; among members of the majority, there is one that positively precludes it.

More important, the account suggests a defense of national partiality against the skeptical argument mentioned above. If certain people have a shared history of doing good, what determines the degree of partiality that is justified between them? Two factors suggest themselves: the degree to which the people's history is shared or involves interaction between them, and the amount of good their interaction produced. Other things being equal, people whose history involves closer relations or more intimate contact have stronger duties of partiality. Also, other things being equal, people whose interactions produced more good, for themselves or for others, have stronger duties of partiality.

The history of family members scores extremely high on the first of these dimensions—namely, closeness of contact. Family members interact intimately on a daily basis, with large effects on each other's lives. Family history also scores high on the dimension of good done, given the large benefits given by parents to their children, spouses to each other, and even children to their parents. Surely family members benefit each other as much as they do any individuals.

A nation's history, by contrast, scores very low on the first dimension. As I have said, I have not met the majority of my fellow Canadians and do not know who they are. But a nation's history does much better on the second dimension. Consider another example from my history. In the 1960s Canadians created a national health care system that continues to provide high-quality medical care to all citizens regardless of their ability to pay. The benefit this medicare system provides any one citizen is probably less than that provided by his or her family, but it is still substantial, and it is one Canadians have provided together. Canadians derive equally substantial benefits from many other aspects of their political activity. When these benefits are added together, they constitute a significant counterweight to the weakness of national relations on the first dimension, that of closeness of contact. The critique of national partiality considers only this first dimension, of closeness. But if we believe that a necessary basis for justified partiality is a shared history, that this history must be good rather than evil, and that the degree of partiality a history justifies depends partly on the quantity of goodness it produces or embodies, we have some response to the critique. On the one dimension, a national history does indeed have much less of the character that justifies partiality than a family history. But on another dimension, the national history has roughly as much.

This account of the basis of national partiality fits most obviously those many nationalisms that point to glorious deeds in the nation's past, such as saving Europe for Christendom or inventing representative democracy. But the account should not be too closely tied to these nationalisms, for two reasons. First, if the basis of

national partiality is objective rather than subjective, it must depend on the nation's actual history rather than on beliefs about that history that are all too often false. A national mythology with no basis in fact cannot justify nationalist policies today.¹⁷ Second, the benefits produced in a nation's history need not be specially grand; on the contrary, they can be perfectly ordinary. Consider again familial partiality. The benefits my wife and I have given each other, such as companionship and love, are also given to each other by countless other couples. What ties my wife and me specially together is not that we have produced unique goods but that we have produced familiar goods jointly, in interactions with each other. The goods in a nation's history can likewise be familiar. Before enacting medicare, Canadians together maintained political institutions and through them the rule of law in Canada, which ensured liberty and security for all Canadian citizens. The same liberty and security were produced in other nations, but only my fellow Canadians produced them with me, and it is that historical fact that is decisive. According to the account I am proposing, it is important that a nation's history have produced significant benefits, but these benefits need not be the grand ones of national mythologies or even at all different from those produced in other nations' histories.

Nations as defined by political institutions¹⁸ are not the only large groups that can have this kind of history. Consider a linguistic and cultural group. Its members have together sustained a language and through it the possibility of beneficial communication for all its speakers. Other groups have also sustained languages, but this group has done it here. They have also, as writers and readers, sustained a literature and an artistic tradition that provide further benefits. When political and cultural groups coincide, these two grounds of partiality reinforce each other. The nation's members have two separate reasons for being partial to the same individuals. But when political and cultural boundaries do not coincide, there can be conflicts about partiality. Consider francophone Quebecers. They share a political

¹⁷ It is also relevant that these beliefs usually concern the distant past, which in my view counts less in justifying partiality than does the recent past. Immediately after World War II, national partiality on the part of Germans would have been morally unthinkable because of the evil their nation had just done. It is much less so today, after fifty years of the Federal Republic.

¹⁸ Those of us who live in multicultural states, especially ones like Canada where two cultures are geographically separated, are much more likely than others to define "nation" in political rather than in ethnic or cultural terms. If we did not, we would be barred by language from any pan-Canadian "nationalism." The Canadian understanding of "nation" is nicely illustrated by an incident from the 1968 federal election campaign. The Progressive Conservatives, seeking to reverse decades of electoral failure in Quebec, announced a *deux nations* policy, according to which Canada was composed of French-speaking and English-speaking *nations*. Pierre Trudeau, recently elected leader of the Liberal party, said he rejected this "two nations" policy. He did not favor the separation of Quebec but wanted Canada to remain one nation. Though Trudeau's reply benefited him electorally in English Canada, it was widely regarded as linguistically mischievous. The English word "nation," it was said, is not equivalent to the French *nation*. While the French word has a primarily cultural significance, the English word is political. The correct translation of *deux nations* is therefore "two [founding] peoples," which does not carry, as "two nations" does, any implication of separate political institutions.

history with all Canadians and a cultural history with a smaller number of franco-phone Canadians. Which group they feel more partial to will depend on how good they think the groups' present qualities are and how beneficial they think the groups' histories have been. Those who think of Canada as a successful country with an admirable political history will be strongly attached to the larger group; those who see present failure and a past of suppressing minorities will not.

Whether a nation is defined politically or culturally, its history differs from a family's in involving many more people, both as recipients of its benefits and as participants in producing them. If only the first of these differences, in the number of beneficiaries, mattered morally, the nation's history would score much higher on the dimension of good done than the family's, since its benefits are much more widely dispersed. The total good resulting from Canadian medicare, for example, is vastly greater than any produced in a family. But it is more plausible to count both differences about numbers, so that what matters for this dimension is not the total benefit produced in a history but something closer to the average benefit per participant, which in the national case roughly equals the average benefit per recipient.¹⁹ Even when we take this view, however, the good produced in a national history is comparable to that in a family history. If we consider the benefits each Canadian receives from living under the rule of law and with social programs such as medicare, they are surely of similar size to those that person receives from his or her family. If this is so, a national history scores roughly as well on the dimension of good done as a family history. Since the national history scores less well on the dimension of interaction, the result on balance is that less partiality is justified toward one's conationals than toward one's family members. This is an intuitively plausible result. Not even the most ardent nationalist claims that one should care as much about one's conationals *as* conationals, as about one's spouse or child. And the degree of concern that is justified toward conationals is considerably greater than toward non-nationals, since one's history with the latter scores very poorly on both dimensions. One not only has had no close interactions with non-nationals but also has produced no significant goods with them. The political and cultural institutions of a nation enable its members to cooperate, however indirectly, in producing significant benefits. But there are no comparable institutions joining non-nationals, even ones living just across a national border, and therefore no comparable goods they can be said jointly to have produced.

I wish I could say more precisely what degree of national partiality this historical account justifies. Unfortunately, that would require weighing against each other more precisely the two dimensions of closeness of contact and good done in

¹⁹ I owe this point to Jeff McMahan. Note that treating both differences as significant plausibly implies that members of small nations have just as strong duties of partiality as members of large nations. If only the number of recipients were significant, citizens of the United States would have a duty of partiality ten times as strong as that of Canadians because their history has benefited ten times as many people.

a history, which I cannot now do. Nor do I see that more precise weightings of these dimensions follow from the general ideas I have advanced. So I will content myself with two more modest conclusions. The first is that, whatever degree of national partiality is intrinsically justified, it is more than the limited degree that the comparison with families initially suggested. Though a national history scores less well on one dimension than a family history, it scores comparably well on another and therefore justifies at least a moderate degree of partiality. It may be that any morally acceptable national partiality must be constrained by respect for the basic rights of all persons, both within one's nation and outside it. But familial partiality is likewise constrained by respect for rights, and it still has considerable room to express itself. The second conclusion is that it is no surprise that nations and cultures are prime objects of partial attitudes. According to the historical account, partiality is justified when the members of a group have worked together in the past to produce significant benefits. But nations and cultures embody just the institutions that make such beneficial interactions possible. My nation is an appropriate object of partial attitudes because it more than other similarly sized groups has allowed me to act with others to produce significant human goods.

Proportionality in the Morality of War

Just war theory, the most widely accepted theory of the morality of war, contains two proportionality conditions that say a war or an act in war is justified only if the damage it causes is not excessive. These conditions have figured prominently in recent debates about the morality of particular wars, including the Gulf, Kosovo, and Iraq wars. But commentators often say the conditions are poorly understood, so it is unclear exactly what they do and do not forbid. In this chapter I will try to clarify the idea of proportionality in war, or explain what makes damage in war excessive. I will not, however, arrive at simple conditions that can be applied mechanically to acts in war. This is partly because judgments about proportionality in war require empirical assessments that are complex and controversial, but also because the conditions themselves can be formulated in different ways that have different implications even given an agreed-on set of facts. I will try to identify some of these differences and show how they affect specific judgments about war. But first I must place the proportionality conditions in the larger context of just war theory as a whole.

I. Proportionality Conditions

Just war theory lays down a series of conditions that a war must satisfy to be morally justified; if it violates any of the conditions it is wrong, although how wrong it

Earlier versions of this chapter were presented to the Rice University Lecture Series on Ethics, Politics, and Society, the University of Calgary, the University of Toronto, a Roundtable on Just War Theory at the University of Illinois/Urbana, and the Murphy Institute Series in Ethical and Political Philosophy at Tulane University. I am especially grateful to my Calgary Ph.D. student David Mellow for conversations on this and related topics over a number of years, as well as to Joseph Boyle, Col. Charles J. Dunlap, Jr., Dennis McKerlie, Jeff McMahan, George Sher, Howard Sobel, Wayne Sumner, the audiences at the above-mentioned occasions, and the Editors of *Philosophy & Public Affairs* for helpful suggestions.

is depends on how many conditions it violates, how important they are, and how seriously it violates them. These conditions are standardly divided into two groups. The *jus ad bellum* conditions concern the resort to war and are directed to political leaders deciding whether to initiate war or whether to respond to another state's doing so with military force of their own. The *jus in bello* conditions concern the means used to fight war. They are again directed at political leaders when they make tactical decisions such as Truman's decision to bomb Hiroshima, but also at soldiers as they fight from day to day. It is usually assumed that the two sets of conditions are independent, so a state can be justified in its resort to war but violate the *in bello* conditions in how it fights, or initiate war unjustly but use only tactics that are morally allowed.

The most important *ad bellum* condition says the resort to war is justified only given a just cause. The most widely accepted just cause for war is resisting aggression, or an armed attack on one's own or another state, but there can also be a just cause when one state sponsors or allows deadly attacks on another's citizens without threatening the other's territory; this was the trigger for the Afghanistan war of 2001. Many theorists now also allow humanitarian just causes, which protect the citizens of another state from rights violations by their own government. Two less important *ad bellum* conditions say a war must be declared by a legitimate authority and fought with right intentions, and three final conditions concern the consequences of war. One says a just war must have a reasonable hope of success; if there is no probability of achieving the just causes, the war's destructiveness will be to no purpose. Another says war must be a last resort; if the just causes can be achieved by less violent means such as diplomacy, fighting is wrong. Last is the *ad bellum* proportionality condition, which says the destructiveness of war must not be out of proportion to the relevant good the war will do. Even if there is a just cause and no way of achieving it other than war, resorting to war can be wrong if the damage it will cause is excessive. For example, the Soviet Union's invasion of Czechoslovakia in 1968 gave NATO a just cause for war, but most people think a military defense of that country would have been horribly wrong because it risked starting a global nuclear war.

There are three *in bello* conditions, of which the first, the discrimination condition, distinguishes between those people who are and those who are not legitimate targets of military force. There is controversy about exactly who these people are, but the traditional view is that deadly force may be directed only at combatants, including soldiers and munitions factory workers, but not at noncombatants. The discrimination condition does not forbid all killing of civilians. It concerns only targeting and therefore allows the killing of noncombatants as a side effect of force directed at properly military targets, or as "collateral damage." In many versions of just war theory, the distinction here turns on the doctrine of double effect, which says it is more objectionable to intend evil as one's end or a means to one's end than merely to foresee that evil will result from what one does. On this reading, the discrimination condition forbids intending the deaths of

noncombatants as an end or means, as in terror bombing that aims to demoralize an enemy by killing its civilians, but does not forbid acts that merely foresee the deaths of noncombatants, as when one bombs an arms factory knowing that some civilians nearby will be killed. Just war theory would be unacceptable if it said there is no objection at all to killing civilians collaterally, but two further conditions prevent this. The necessity condition, which parallels the *ad bellum* last-resort condition, says that killing soldiers and especially civilians is forbidden if it serves no military purpose; unnecessary force is wrong. And the *in bello* proportionality condition says the collateral killing of civilians is forbidden if the resulting civilian deaths are out of proportion to the relevant good one's act will do; excessive force is wrong. This proportionality condition is included in Additional Protocol I to the Geneva Conventions, which forbids attacks "which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."¹ The condition allows bombing a vital munitions factory if that will unavoidably kill a few civilians, but forbids killing thousands of civilians as a side effect of achieving some trivial military goal.

The proportionality conditions are actually more important than this initial account suggests, since, if formulated properly, they can incorporate the other just war conditions about consequences. Imagine that a war has no chance of achieving any relevant goods. This fact, which makes it violate the reasonable hope of success condition, surely also makes it disproportionate, since its destructiveness now serves no purpose whatever. The same is true if the war has only some small probability of achieving relevant goods, since then its expected harm is excessive compared to its expected good. If it takes account of probabilities in this way, as on any plausible view it must, the *ad bellum* proportionality condition incorporates hope-of-success considerations, and it can also incorporate last-resort considerations. Now imagine that a war will achieve certain goods at not too great a cost, but that the same goods could be achieved by diplomacy. Here the war may not be disproportionate in itself, but it is disproportionate compared to the alternative, since it causes additional destruction for no additional benefit. The same is true in the more realistic case where the war will achieve some minor relevant goods to a slightly higher degree. Here too it is disproportionate compared to diplomacy, since it imposes significant additional harms for the sake of insignificant benefits. Michael Walzer has said that, if taken literally, the last resort condition would make war

¹ 1977 *Geneva Protocol I Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts*, Art. 51 (5) (b), in Adam Roberts and Richard Guelff, eds., *Documents on the Laws of War*, 3d ed. (Oxford: Oxford University Press, 2000), p. 449. This Protocol has not been ratified by the United States, but its military accepts very similar statements. Thus, the United States Army's Field Manual 27-10 says, about the bombing of defended places, that "loss of life and damage to property incidental to attacks must not be excessive in relation to the concrete and direct military advantage to be gained" (U.S. Department of the Army, *The Law of Land Warfare* [Field Manual 27-10, 1956], par. 41).

morally impossible, since “we can never reach lastness . . . there is always something more to do: another diplomatic note, another UN resolution, another meeting.”² For precisely this reason the condition must not be read literally, but must assess the alternatives to war in the same way as it does war: for the relevant good they may produce, their probability of producing it, and any costs that will result if the alternatives are tried and fail, such as making an eventual war more bloody. But then the last resort condition is in effect a comparative version of the initial, simple proportionality condition. For war and each of its alternatives it does a proportionality calculation, identifying the relevant goods and evils it will produce compared to a baseline of doing nothing, or continuing to act as one would have had there been no just cause.³ This yields the net good or bad effects of each, and it then says war is permitted only if its net outcome is better than those of all alternatives. If the *ad bellum* proportionality condition is extended to make these comparative judgments, it incorporates last-resort considerations, and the *in bello* condition can likewise incorporate necessity considerations if it compares the net effects of a particular tactic in war with those of alternatives that may be less harmful.

Each proportionality condition allows two formulations. An objective version assesses a war or act in light of its actual effects, that is, the relevant good it actually produces and its actual destructiveness; a subjective version does so considering only an item’s likely effects given the evidence available to agents at the time. Both versions must make some probability estimates: of the likely effects of alternatives that are not chosen (for comparative conditions like last resort) and of the magnitude of evils the war does good by preventing. But given their different assumptions about a war’s positive effects the two can yield different results, so a war can be objectively proportionate but subjectively disproportionate, or vice versa.

Despite their differences, the various proportionality conditions—*ad bellum* and *in bello*, simple and comparative, objective and subjective—all say a war or act in war is wrong if the relevant harm it will cause is out of proportion to its relevant good. This raises three questions: (1) What are the relevant goods that count in favor of a war’s or act’s proportionality? (2) What are the relevant evils that count against it? (3) How do these goods and evils weigh against each other? I will begin with the first question, about goods. But first a more general comment is in order.

As many writers have noted, the structure of just war theory closely parallels that of the morality of self-defense.⁴ The latter too allows the use of force only for certain ends, namely to protect one’s own or another’s rights, and limits that force by proportionality and necessity conditions. An act of self-defense is wrong if the harm it causes the attacker is out of proportion to the harm he threatens, or if the threat could just as well have been averted by less violent means. These parallels

² Michael Walzer, *Just and Unjust Wars*, 2d ed. (New York: Basic Books, 1992), p. xiv.

³ For a lucid defense of this “do-nothing” baseline for proportionality judgments, see David Mellow, “A Critique of Just War Theory” (Ph.D. diss., University of Calgary, 2003).

⁴ See, e.g., the discussion of the “domestic analogy” in Walzer, *Just and Unjust Wars*, pp. 58–59.

suggest a promising line of argument. Since our intuitions about self-defense are often clearer than our intuitions about war, we can try to make progress with just war proportionality by considering parallel cases involving individuals: if we think certain uses of force are not permitted by individuals, we can reach parallel conclusions about force in war. This type of argument cannot be decisive, since there is no guarantee that what holds in the two domains is identical. But it is suggestive, and I will use it in that spirit.

II. Relevant Goods

The simplest view of proportionality in war is a quasi-consequentialist one that counts all the goods and evils that result from a war or act in war and weighs them equally, so a choice is disproportionate if the total evil it causes is greater than its total good. James Turner Johnson defends this view about *ad bellum* proportionality, saying it requires the “total good” caused by war to outweigh the “total evil,” or that “the overall good achieved by the use of force...be greater than the harm done.”⁵ His view does not result in a completely consequentialist theory of war, for two reasons. First, even a war with overall optimal effects can be wrong if it violates other just war conditions, for example, by lacking a just cause. Second, his view does not require a war to have overall optimal effects, only ones that involve more good than evil. But Johnson’s view does have consequentialist elements, since it counts all the goods and evils a war produces and weighs them equally against each other.

A related view retains the first of these elements but weighs good effects somewhat more heavily than bad ones, so a war can be proportionate even if it causes somewhat more harm than good. The United States Catholic bishops may take this line when they formulate *ad bellum* proportionality as saying, “the damage to be inflicted and the costs to be incurred must be proportionate to the good expected by taking up arms.”⁶ By speaking simply of “costs” and “good expected” they seem to count all resulting goods and evils, but in requiring the evils only to be proportionate to, rather than no greater than, the goods, they may allow the goods to be somewhat smaller. This more permissive view is explicitly defended by Douglas Lackey, who likewise counts all resulting goods and evils but says it would be “too restrictive” to weigh them equally and concludes that “a war for a just cause passes the test of proportionality unless it produces a *great deal* more harm than good.”⁷

⁵ James Turner Johnson, *Morality and Contemporary Warfare* (New Haven: Yale University Press, 1999), pp. 27–28; see also his *Just War Tradition and the Restraint of War* (Princeton, NJ: Princeton University Press, 1981), p. 204, and *Can Modern War Be Just?* (New Haven: Yale University Press, 1984), pp. 3, 25, 62.

⁶ U.S. Catholic Bishops, *The Challenge of Peace: God’s Promise and Our Response*, reprinted in Jean Bethke Elshtain, ed., *Just War Theory* (New York: New York University Press, 1992), p. 101.

⁷ Douglas P. Lackey, *The Ethics of War and Peace* (Englewood Cliffs, NJ: Prentice-Hall, 1989), pp. 40–41.

I think both views are wrong to count all the goods a war will produce. Imagine that our nation has a just cause for war but is also in an economic recession, and that fighting the war will lift both our and the world's economies out of this recession, as World War II ended the depression of the 1930s. Although the economic benefits of war here are real, they surely cannot count toward its proportionality or make an otherwise disproportionate conflict proportionate. Killing cannot be justified by merely economic goods, and the same is true of many other goods. A war may boost scientific research and thereby speed the development of technologies such as nuclear power; it may also satisfy the desires of soldiers tired of training and eager for real combat. Neither of these goods seems relevant to proportionality or able to justify killing; an otherwise disproportionate war cannot become permissible because it has these effects.

Some restriction is needed on the goods that count toward proportionality, and it seems obvious what it should be: the relevant goods are only those contained in the just causes. If a war has certain just aims, the goods involved in achieving those aims count toward its proportionality but goods incidental to them, such as boosting the economy or science, do not. This restriction is included, even if implicitly, in many traditional formulations of *ad bellum* proportionality, which equate the just cause with the prevention of some injury and say the destructiveness of war must not be excessive compared to that injury. Thus, Joseph C. McKenna says that *ad bellum* proportionality requires "the seriousness of the injury [to] be proportionate to the damages that the war will cause," while Richard J. Regan calls the resort to war justified only if "the wrong to be prevented equals or surpasses the reasonably anticipated human and material costs of the war."⁸ These statements concern simple proportionality, but a similar point applies to comparative proportionality, or the last resort condition. It too weighs the destructiveness of war and its alternatives against only their contribution to the just causes, and therefore counts as relevant only alternatives that pursue those causes. This is why the condition does not require a war to have overall optimal effects: a war can be proportionate even if it produces less good than some alternative that does not achieve the just causes. Thus, the Gulf War would have been disproportionate if there had been some less destructive way of evicting Iraq from Kuwait, but not simply if the money it cost would have done more good if spent on development aid to Africa.

In formulating this restriction we should distinguish, as traditional formulations do not, between two types of just cause that Jeff McMahan and Robert McKim call "sufficient" and "contributing" just causes.⁹ Sufficient just causes suffice by themselves to fulfill the just cause condition; they include resisting aggression

⁸ Joseph C. McKenna, "Ethics and War: A Catholic View," *American Political Science Review* 54 (1960): 651; Richard J. Regan, *Just War: Principles and Cases* (Washington, D.C.: Catholic University of America Press, 1996), p. 63. The need for this restriction is explicitly defended in Jeff McMahan and Robert McKim, "The Just War and the Gulf War," *Canadian Journal of Philosophy* 23 (1993): 512–13.

⁹ McMahan and McKim, "The Just War and the Gulf War," pp. 502–06.

and preventing major humanitarian wrongs such as genocide and ethnic cleansing. Contributing just causes do not suffice to satisfy the just cause condition; given only these causes, one is not permitted to fight. But once there is a sufficient just cause, contributing causes can be further legitimate aims in war and can contribute to its justification. They include disarming an enemy and thereby incapacitating it for future aggression, as well as deterring aggression by showing this nation and others that aggression does not pay. If incapacitation is only a contributing cause, the fact that an enemy has armaments it may use aggressively in the future is not a sufficient justification for military action; there is no such far-reaching right of pre-emption. On this view Israel's 1981 raid on Iraq's Osirak reactor was wrong. But once Iraq supplied a sufficient just cause for war by invading Kuwait, disarming Iraq became a legitimate goal. The coalition forces were permitted to pursue that goal into Iraqi territory after liberating Kuwait, and also to include conditions about disarmament in the ceasefire agreement that ended the war. The benefits of incapacitation and deterrence also count toward the war's proportionality, and wars that are not proportionate in themselves, such as perhaps the Falklands War, can become so given their effects on international security. The situation exactly parallels that of criminal punishment. If a person has not yet committed a crime, the fact that he may or even is likely to do so in the future is not a sufficient ground for imprisoning him now. Once he has committed a crime, however, incapacitating him from committing further crimes and deterring other would-be criminals become legitimate aims of punishment, and his sentence can be adjusted to better achieve them.¹⁰ The same holds for other contributing causes, such as lesser humanitarian aims. I think most people would say that the Taliban's repression of Afghan women was not a sufficient just cause; a war fought only to end that repression would have been wrong. But once there was a sufficient just cause in the Taliban's harboring of terrorists, the fact that the war would improve the lot of Afghan women became a factor that counted in its favor and helped make it proportionate.¹¹ There can also be economic contributing causes. Imagine that a war will not give the world's economy a boost but will prevent it from being harmed by being pushed into recession. If the harm will come from another country's exercise of its legal rights, such as shutting off its own oil exports, preventing that harm is neither a sufficient nor a contributing just cause. But imagine that in 1990 Iraq had occupied both Kuwait and Saudi Arabia and stopped all their oil production. In that case the economic harm would have resulted from an unjust act of aggression, and preventing it would have been a relevant benefit of war.

¹⁰ I borrow this analogy from McMahan and McKim, "The Just War and the Gulf War," pp. 504–05.

¹¹ It may be argued that lesser humanitarian aims are in fact sufficient just causes, just not ones that on their own satisfy proportionality; this is especially plausible if we think of the just cause condition as specifying *types* of goal that can justify war, whether or not every instance of them does. Even so, the point remains that goals that would not by themselves justify war can contribute to a war's proportionality when joined with other, weightier goals.

In assessing these contributing causes we must continue to compare them with the do-nothing baseline of having a just cause but not pursuing it. This is vital because often acquiescing in aggression not only fails to deter future aggression but positively encourages it, by allowing a precedent of successful aggression. In the lead-up to the Gulf War many commentators called for a negotiated Iraqi withdrawal from Kuwait, but it was evident that any such outcome would require concessions to Iraq, for example, about some disputed islands on the Iraq-Kuwait border.¹² And these concessions would have encouraged future aggression by showing that one can invade a neighbor and come out ahead. The United States and its closest allies strenuously resisted this approach, insisting that there be “no rewards for aggression.” Whatever its overall merits, their stance recognized that once aggression has occurred, the status quo before the aggression may no longer be an option. One can resist the aggression, which will deter future aggression, or not resist, which will encourage it, and the benefits of the first choice must include avoiding the harms of the second.

The view I am proposing about *ad bellum* proportionality is intermediate between the quasi-consequentialist and traditional views. It does not count all the goods a war will produce nor only those in the sufficient just causes, but supplements the latter with a finite number of contributing just causes. This raises the question of whether there is some unifying feature that gives these contributing causes their status. So far as I can see, there is not; like the sufficient just causes, they are merely the items on a list. But there are intuitive limits on what can go on this list. Shortly after its end, it looked as if the Gulf War would help resolve the Israeli-Palestinian conflict, through the Oslo Accords it made possible. The effects here were of the right kind to be contributing causes, since they involved preventing violence and reducing international tension. But I do not think that, even if realized, they would have counted toward the war’s proportionality, because they would not have been connected to it in the right way. They would not have resulted directly from the war’s sufficient just causes but would have been side effects of the process of achieving them, namely the building of a UN-sponsored coalition combining Western and Arab states. But similar effects that do arise directly from a war’s sufficient just causes can count. If the Iraq War, by eliminating Iraq’s payments to the families of Palestinian suicide bombers, had reduced Palestinian terrorism and thereby encouraged an Israeli-Palestinian settlement, that would have been relevant to the war’s proportionality because eliminating support for terrorism is a legitimate aim.

If *ad bellum* proportionality counts only the goods in a war’s just causes, something similar must be true of *in bello* proportionality. When a particular act in war is justified it is primarily because it contributes to the just causes, by

¹² Lawrence Freedman and Efraim Karsh, *The Gulf War 1990–91: Diplomacy and War in the New World Order* (London: Faber & Faber, 1993), chs. 11, 19.

increasing either the likelihood or the degree of their achievement. That the act will boost scientific research by testing some new weapons system or please soldiers who enjoy testing weapons is irrelevant. Here again contributing just causes must be counted. If disarming an aggressor is a legitimate war aim, particular acts leading to disarmament will be proportionate only if that contributing cause is relevant; the same goes for lesser humanitarian aims. But in the *in bello* case there is another consideration. Sometimes an act that will foreseeably kill more civilians than some alternative will, while not contributing more to the just causes, reduce either the number of our military casualties or our economic costs, perhaps because it uses less expensive and therefore less accurate weapons. Often this fact will not make the act proportionate; we must accept the greater casualties or expense. But the demand here cannot be unlimited: we cannot be required to sacrifice hundreds of soldiers or spend billions of dollars to save a few enemy civilians. If so, *in bello* proportionality must consider as relevant goods an act's contribution not only to the war's just causes but also to reducing the costs of achieving them.

If *in bello* proportionality looks even partly at the just causes for war, it cannot be assessed independently of *ad bellum* considerations, and especially of the moral importance of those causes. Intuitively this seems right. The level of destruction permitted in a war against a genocidal enemy such as Nazi Germany is surely greater than in the Falklands War. But this claim contradicts the dominant view in the just war tradition, which treats the *jus in bello* as entirely independent of the *jus ad bellum*, so the same *in bello* rules apply to both sides of a conflict whatever the justice of their aims. This independence is affirmed in the Preamble to Additional Protocol I, which says its provisions apply to all persons "without any adverse distinction based on the nature or origin of the armed conflict or on the causes espoused or attributed to the Parties to the conflicts." It is also reflected in the Protocol's statement of *in bello* proportionality, which says damage to civilians must not be excessive "in relation to the direct military advantage anticipated," with no reference to the further goods such advantage will promote.¹³ Now, the independence of *in bello* considerations is plausible for the discrimination condition as traditionally understood, since, whatever their war aims, both sides can refrain from targeting noncombatants. But it does not fit the *in bello* conditions about consequences, namely proportionality and necessity. If "military advantage" justifies killing civilians, it does so only because of the further goods such advantage will lead to, and how much it justifies depends on what those goods are. Compelling though it is, this view has the radical implication that no act by soldiers on a side without a just cause can satisfy proportionality: if their acts produce no relevant goods, they can never be just. It does not follow that these soldiers should be punished. Since soldiers normally cannot be expected to evaluate their

¹³ *Additional Protocol I*, Preamble and Art. 51 (5) (b), in Roberts and Guelff, *Documents on the Laws of War*, pp. 422–23, 449.

nations' war aims, they are not to blame for acting wrongly, and international law is probably best formulated as it is, with no "adverse distinction" against soldiers on an unjust side. The alternative approach might even lead to more destructive wars since, notoriously, often both sides in a conflict believe their cause is just.¹⁴ Nonetheless, if we consider the morality of war rather than its legality, the independence of its two branches cannot be maintained. Whether an act in war is *in bello* proportionate depends on the relevant good it does, which in turn depends on its *ad bellum* just causes.¹⁵

III. Relevant Evils

When we turn to the evils relevant to proportionality, we seem to find no restriction on their content parallel to the one on relevant goods. That a war will boost the world's economy does not count in its favor, but that it will harm the economy surely counts against it. Whereas economic benefits are not relevant goods for proportionality, economic harms are relevant evils. It is also relevant that a war will hamper scientific research or cause pain to the soldiers who fight; these effects too can make a war disproportionate. In assessing a war for proportionality, it seems we count evils of all the kinds it will cause, with no limits on their content. There is therefore a thumb pressed down on one side of the proportionality scale, with more counting on the negative than on the positive side. But there may be another, compensating thumb on the positive side.

Although restricted in their content, the goods relevant to proportionality seem not to be restricted by their remoteness from a war or act either in time or causally. If defending a nation will ensure that its citizens are self-determining a century from now, that helps to make the defense proportionate, and the same holds if defending them will deter wars a century from now. If some war really would end all war forever, that would count massively in its favor, and it would do so even if the good effects required intervening good decisions by other agents, namely decisions not to go to war. The same view can in principle be taken of relevant evils. One objection raised before both the Gulf and Iraq wars was that they would increase instability in the Middle East and so lead to further violence in the region. This objection counts encouraging war as a relevant evil and seems to do so without

¹⁴ This worry is expressed in Judith Gail Gardam, "Proportionality and Force in International Law," *American Journal of International Law* 87 (1993): 392–94.

¹⁵ The most prominent recent defender of the independence of the *jus ad bellum* and *jus in bello* has been Walzer, but his *Just and Unjust Wars* also downplays the importance of the proportionality conditions. These features of his view may be connected. While the discrimination condition Walzer concentrates on is indeed, on his reading, independent of the *jus ad bellum*, the *in bello* proportionality condition depends on it. More general attacks on the independence of the *jus in bello*, extending to the discrimination condition, are mounted in Jeff McMahan, "Innocence, Self-Defense and Killing in War," *Journal of Political Philosophy* 2 (1994): 193–221, and Mellow, "A Critique of Just War Theory."

any restriction about remoteness. But the objection is much more persuasive for third-party interveners than for a nation that is itself under attack. Imagine that we face aggression from our eastern neighbor. We may know that if we defend ourselves successfully this will encourage our western neighbor to expand its military, once it sees how useful a military can be. And that will cause its further western neighbor, with whom it has various disputes, to expand its military, leading to an arms race and eventually to war between them. Now, war between our neighbors is the right type of evil for proportionality calculations, but it is not clear that it is one we must weigh heavily in our decision or that can take away our right of national defense. Surely we would not be condemned if, despite knowing the consequences, we chose to resist the aggression. The parallel certainly holds in individual self-defense. If my defending myself against an attacker will lead to increased attacks on others, perhaps because my attacker will vent his frustration on weaker victims, this fact, though regrettable, does not make my defending myself wrong. Many will say the reason is that, even if the later attacks would not have occurred but for my defense, the responsibility for them belongs not to me but to the attacker who launches them. Similarly, in the international case many will say that if defending ourselves from our eastern neighbor leads to war between our western neighbors, that is their responsibility, not ours. Although the right type of effect for proportionality, that war is too remote from our choice to weigh heavily against it.

The suggestion here is that the intervention of another's wrongful choice can reduce our responsibility for a bad outcome, and in fact this possibility arises frequently in war. Imagine that we have a just cause but know that if we pursue it some fanatics on the other side will with no moral excuse launch suicide attacks on our civilians. In deciding whether to fight we may have to weigh the potential deaths of our civilians, but what about those of the suicide bombers? Is the fact that they will die in unjust attacks one that could make our choice of war disproportionate? That is hard to believe. Nor must the relevant wrong choice come after our act. Imagine that in the same war enemy civilians with likewise no moral excuse act as voluntary shields, placing themselves around the enemy's military installations in the hope of deterring us from attacking those installations. Here the fact that our attack will foreseeably kill noncombatants is not morally irrelevant; if we have a choice between attacking two installations of equal military importance, only one of which has shields, we should prefer attacking the other. But by placing themselves near the installations the shields arguably take upon themselves some responsibility for their deaths and remove it from us, so their deaths count less against our attack's proportionality.

In these examples the intervening choices are by the individuals who will suffer the harms, but this too is not necessary. Consider, first, involuntary shields, who are placed near military targets against their will or, more commonly, have military personnel and equipment located in their neighborhoods, again in the hope of deterring attack. Additional Protocol I forbids all use of civilians as shields but

immediately adds that the violation by one party of its obligations under the Protocol does not release other parties from their obligations, suggesting that the protections for involuntary shields are in no way reduced.¹⁶ Not everyone takes this view, however. Discussing the Vietnam War, in which Viet Cong guerrillas hid among the Vietnamese peasantry, Paul Ramsey says, “the onus for having placed multitudes of peasants within range” of fire belongs to the guerrillas and not the United States, since “[t]o draw any other conclusion would be like, at the nuclear level, granting an enemy immunity from attack because he had the shrewdness to locate his missile bases in the heart of his cities.”¹⁷ William V. O’Brien concurs: “It seems fair to assign the major responsibility to the Communist forces for the civilian losses, destruction, and displacement caused by turning the population centers into battlefields.”¹⁸ A similar view may have been taken by the U.S. military in the Iraq War. Early in that war a fight outside Nasiriyah moved into the city when Iraqi forces retreated there, with resulting civilian casualties. The commander of a U.S. artillery battalion firing on Nasiriyah “placed responsibility for any civilian deaths on the Iraqi soldiers who drew the marines into the populated areas,” saying, “We will engage the enemy wherever he is.”¹⁹

Or consider the coalition bombing campaign in the Gulf War. It initially caused around 2,000 Iraqi civilian deaths, but many more followed from its aftereffects, especially the damage to Iraq’s water filtration plants. Some of these later deaths were unavoidable, but others could have been prevented had the Iraqi government repaired the country’s infrastructure more quickly, as it arguably had a moral duty to do. In assessing the bombing for proportionality, then, do we count all the civilian deaths that resulted given the Iraqi government’s actual behavior, or only those that would have resulted had that government acted as it should? A similar issue arises about the economic sanctions that followed the war. Critics say they caused the deaths of 500,000 Iraqi children; defenders reply that the deaths were Saddam Hussein’s responsibility rather than the UN’s, since he could have prevented many of them by making fuller use of the UN’s oil-for-food program, and could have prevented all of them by openly abandoning his pursuit of weapons of mass destruction, as he had promised in the ceasefire agreement to do. Since these failures of his were necessary for the sanctions to cause the deaths, those deaths are properly charged to him and not to the sanctioning nations.²⁰

¹⁶ *Additional Protocol I*, Art. 51, (7–8), in Roberts and Guelff, *Documents on the Laws of War*, p. 449.

¹⁷ Paul Ramsey, *The Just War: Force and Political Responsibility* (New York: Scribner’s, 1968), p. 437.

¹⁸ William V. O’Brien, *The Conduct of a Just and Limited War* (New York: Praeger, 1981), p. 100. Johnson takes a similar view of Israel’s response to the PLO in the early 1980s; see *Can Modern War Be Just?* pp. 57, 59. I take it there is no point in assigning the “major” responsibility for civilian deaths to the side that uses them as involuntary shields unless this reduces the other side’s responsibility.

¹⁹ “Marines Wade into Dreaded Urban Battle,” *The Globe and Mail*, March 25, 2003.

²⁰ Kenneth Pollack, *The Threatening Storm: The Case for Invading Iraq* (New York: Random House, 2002), pp. 133–37, 139–40; Marcus Gee, “Stick to your Poetry and Pianos,” *The Globe and Mail*, Sept. 27, 2002, p. A17; and Margaret Wentz, “Who’s Killing the Children of Iraq?” *The Globe and Mail*, Oct. 8, 2002, p. A23.

The question of whether others' wrongful choices can reduce our responsibility for bad outcomes is vital for the analysis of just war proportionality, but it is very difficult to answer decisively. One extreme view says another's wrong choice always completely removes our responsibility for resulting evils, but this in effect eliminates proportionality as an independent just war condition. Any time we have a just cause, an enemy's resisting us is wrong, and any evils that follow from his resisting, namely all the evils of the war, are his responsibility, not ours. On this view, if NATO had challenged the Soviet invasion of Czechoslovakia in 1968 despite knowing that nuclear war would result, there would have been nothing objectionable about its choice. That is very hard to accept.²¹ A contrary extreme view ignores others' agency completely and counts all resulting evils, whatever other causal conditions they may have. It is not so decisively objectionable and may even be correct, but it is at least questionable when it gives the deaths of suicide bombers and voluntary shields completely unreduced weight. And there are further views intermediate between these extremes.

First, a view can give evils that depend on others' agency only diminished rather than zero weight, holding only that others' cooperation in producing a bad outcome reduces our responsibility for it somewhat. Second, a view can make this reduction only in some cases and not others. One possibility is to discount evils only when the intervening choice is by the very person who will suffer the evil; this view discounts for suicide bombers and voluntary shields but in few other cases.²² This view may again be correct, but it will not be attractive to those who want to discount even a little for involuntary shields, when an enemy fails to rebuild after a bombing, or for economic sanctions. And there are several ways of capturing these cases short of adopting the extreme view that discounts for all intervening agency. For example, a view can reduce our responsibility for resulting evils only when the intervening choice either (1) is by the person who will suffer the evil, or (2) only affects the amount of harm *our* act will cause without introducing new causal processes leading to new harms.²³ This view's second clause applies to the three cases just mentioned, since in all of them the enemy's choice affects only how much harm our bombing or sanctions cause. But it does not apply to the case of

²¹ Paul Christopher comes close to adopting this view. To the charge that the first Gulf War was disproportionate because it caused the deaths of 40,000 Iraqi soldiers, he replies that "the responsibility for the deaths of Iraqi soldiers must rest with the Iraqi government," since "Iraq could have prevented or ended the war at any time by complying with the mandate of the United Nations Security Council" (*The Ethics of War and Peace: An Introduction to Legal and Moral Issues*, 2d ed. [Upper Saddle River, NJ: Prentice-Hall, 1999], pp. 84–85).

²² The intuitive conclusion about suicide bombers and voluntary shields can also be reached in another way, by a view that discounts all deaths on the enemy side by their victims' moral culpability with respect to the war or to acts in it. But this view raises issues beyond the scope of this article and departs, as I am trying not to, from the assumptions about legitimate targets found in international law.

²³ The first disjunct is needed because the second disjunct, while discounting for voluntary shields, does not do so for suicide bombers.

Czechoslovakia in 1968, since then the nuclear war would have resulted from a Soviet military response to NATO that created new harmful processes. There is in fact a whole range of intermediate views about intervening agency, depending on how much they reduce responsibility (a lot or only a little) and in how many cases (almost all or only a few). It is difficult to choose between these views, as well as between them and the simpler view that counts all resulting evils. I cannot find any abstract argument that favors one view over the rest, nor do intuitive judgments about particular cases yield a decisive result. On the contrary, there are sharp disagreements about, for example, the U.S. treatment of Vietnamese peasants and the sanctions against Iraq. I will therefore have to leave this issue unresolved, despite its vital importance for just war proportionality. The more the proportionality conditions discount resulting evils for others' wrongful agency, the more permissive those conditions are; the less the conditions discount, the more wars and acts they forbid. On the one view, the sanctions against Iraq may have been entirely unobjectionable; on the other, grossly disproportionate.

IV. Weighing Goods and Evils: Defending Sovereignty

Having identified their relevant goods and evils, the proportionality conditions must then weigh them against each other. To isolate the distinctive issues here I will assume that resulting evils are not discounted for others' agency, but instead count fully against relevant goods. Even so, the task of weighing is complex, with as many aspects as there are pairs of relevant goods and evils. Thus, a full account of just war proportionality must weigh the defense of a nation's sovereignty against environmental harms, the benefit of liberating women against economic costs, and more. I cannot discuss all these comparisons, but will instead focus on two issues I take to be central: whether defending sovereignty against aggression ever justifies killing, and how a nation should weigh its own citizens' lives against lives on the other side.

Resisting aggression is usually considered the paradigm just cause for war and one that almost always satisfies proportionality. Immediately after introducing the *ad bellum* condition, McKenna adds that "Self-defense...almost always justifies resistance."²⁴ On this view there may be special circumstances, such as those of Czechoslovakia in 1968, where the consequences of national self-defense are so catastrophic as to make it wrong, but where resisting aggression will lead only to conventional war it is normally proportional. This view makes sense given a traditional understanding of just war theory, on which the entities with rights in the international realm are states, understood as indivisible entities with a status parallel to that of individuals in the morality of

²⁴ McKenna, "Ethics and War," p. 651.

self-defense. Then a state facing aggression faces a threat to its existence, just as an individual does whose life is attacked; and just as the individual may kill an attacker to protect his existence, so may the state kill. But this “statist” view has been persuasively criticized on the ground that all rights belong ultimately to individuals. It is individuals who at bottom matter morally, and any rights states have must derive from and concern the rights of their citizens.²⁵ This “individualist” view has important practical implications, but it also tightens the relation between just war theory and the morality of self-defense, making the former not just parallel to but derivative from the latter. If the only just cause for war is to protect the rights of individuals, then legitimate military action always is an instance of defending individuals. And if the state acts legitimately only when it acts on authority given it by its citizens, as many liberal theories hold, then any limitations on their enforcement rights must extend to its own. The morality of individual self-defense permits a person to defend not only himself or one other person but also a group; if a hundred people are attacked on the street, he may try to defend them all. It also permits people to coordinate their defensive acts, so a hundred act jointly to defend one. The individualist view makes it natural to see legitimate military action as extending these two possibilities, so in it a large group of individuals act collectively, through their political institutions, to protect the rights of another large group of individuals, who may be themselves.²⁶ Then cases of individual defense are not just analogous to cases in war; they concern the same topic.

The most-noticed practical implication of the individualist view has been for humanitarian intervention. Whereas the statist view forbids armed interference by one state in the internal affairs of another, the individualist alternative allows such intervention to prevent serious violations of citizens’ rights by their own government, as in Rwanda or Kosovo. But this view also has implications for national defense, which it says satisfies proportionality only if it protects rights of citizens that are important enough to justify killing. This condition is satisfied if the aggressor plans genocide or other serious crimes against the nation’s citizens, such as rape, but often an aggressor has no such aim. It seeks only to absorb the nation’s territory and replace its government, changing the citizens’ political status but not much else about their lives. The aggressor will kill the nation’s citizens if they resist its attack, but not if they do not. The unavoidable threat it poses is therefore only to the citizens’ political rights, such as their rights to participate in collective political self-determination. Two writers, Richard Norman and David Rodin, have argued that these rights are not important enough to justify killing. If someone tries to prevent me from voting, for example, I am not permitted to kill him in response. Norman and Rodin conclude that if a nation faces aggression that

²⁵ See, e.g., David Luban, “Just War and Human Rights,” *Philosophy & Public Affairs* 9 (1979/80): 160–81.

²⁶ I owe this last argument to Jeff McMahan.

does not threaten further rights of its citizens, it is not permitted to defend itself with lethal military force.²⁷

Some may find this a philosophers' argument in the pejorative sense, one completely divorced from reality. In the real world of international relations there is no debate about the right of national self-defense, which is firmly entrenched in the UN Charter. But if this consensus is correct it must be possible to show why, and I will now attempt this. While granting that the rights relevant to *ad bellum* proportionality belong only to individuals, I will argue that they have three features that differentiate the situation of a nation facing aggression from that of an individual protecting his right to vote. I will not claim that any of these features alone justifies killing, only that they do so together.

The first and most obvious feature is the number of people whose rights are threatened. Whereas the voting example involves just one person, aggression threatens the political rights of all a nation's citizens, who can number tens or even hundreds of millions. In addition, if resisting the threat will deter future aggression, it protects the rights of many other nations' citizens, and in each case the protection is for an extended time, since a successful invasion violates rights not momentarily but for many years. The question then is whether this factor of greater numbers affects the amount of defensive force a victim is permitted to use. In some aspects of the morality of self-defense it does not. If a person is attacked by a group of aggressors and can save his life only by killing them all, he is permitted to do so no matter how large the group. So numbers do not count on the side of aggressors, and in some cases they do not make a difference on the side of victims. Not only may one person not kill to prevent himself from being tickled, but a group may not kill to stop themselves all from being tickled. Not even a million people may kill to save themselves from that trivial a threat.²⁸ But it seems that in other cases the number of victims does make a difference. More specifically, I think the number of victims can boost the amount of force permitted in response to a threat to some degree even if not always to the point of killing. Thus, although one person is not permitted to break an assailant's arm to prevent himself from being tickled, a large enough group may be permitted to break an arm to prevent themselves from being tickled. Similarly for duration: while a person may not be permitted to break an arm to prevent himself from being confined in a room for five minutes, he is surely permitted to do that and more to prevent himself from being confined for fifty years. Parallel claims are certainly plausible for other cases in war: surely more force is permitted to prevent 100,000 Kosovar Albanians from being expelled from

²⁷ Richard Norman, *Ethics, Killing and War* (Cambridge: Cambridge University Press, 1995), ch. 4; David Rodin, *War and Self-Defense* (Oxford: Clarendon Press, 2002), ch. 6.

²⁸ In F. M. Kamm's terminology, the benefit of not being tickled is here an "irrelevant utility"; in Samantha Brennan's, it does not satisfy the "universal constraint" a benefit must satisfy if it is to help justify killing. See Kamm, *Morality, Mortality, Vol. I: Death and Whom to Save from It* (New York: Oxford University Press, 1993), p. 146; and Brennan, "Thresholds for Rights," *Southern Journal of Philosophy* 32 (1995): 149–53.

their homes than to prevent one from being expelled. So while granting that there is a limit on the extra defensive force a large number of victims permits against a threat, one can hold that it boosts that force somewhat. When a nation faces aggression, the threat is to an immense number of people's rights for an immense period of time. Even if this does not by itself justify killing, it justifies more force than is permitted to protect one person's one-time exercise of his right to vote.

Second, a military aggressor backs up its attack on another nation with the threat to kill its citizens if they resist. Although no citizens will be harmed if they surrender, they will be killed if they do not. McMahan has argued that this conditional threat by itself licenses a victim to kill in self-defense: by threatening one's life an attacker brings one's right to life into play and permits whatever degree of force defense of that right allows.²⁹ Norman and Rodin reply persuasively that this is not so: if a thief threatens to kill you if you do not give him a dollar, you may not kill him to protect your dollar. But it does not follow, as Norman and Rodin seem to assume, that the threat has no effect on the amount of force one may use; on the contrary, the threat again seems to boost this amount somewhat. A victim is presumably not permitted to break a thief's arm to prevent him from taking a dollar, but he may be permitted to do so if the thief backs up his attempt with the threat to kill. It is instructive to consider Norman's and Rodin's analysis of this case. Both say the thief's action has two components: the direct attempt on the victim's dollar and the conditional threat to the victim's life should he resist that attempt. The first element on its own does not justify killing, since that would be a disproportionate response to a threat to one's dollar. Nor does the second element, since killing is unnecessary when one can escape the threat by handing over the dollar. If neither component on its own justifies killing, Norman and Rodin conclude, the thief's action as a whole cannot.³⁰ But this analysis implies that from the thief's point of view threatening the victim's life provides benefits at no cost, since it increases his chance of getting the dollar while not increasing the force his victim may use against him. One does not have to be a consequentialist to find this troubling. I think it is more plausible to see an attempt at a victim's dollar backed by a threat to kill as a single violation that is more serious than a simple attempt at the dollar and less serious than a direct attempt to kill, so it justifies more defensive force than the former but less than the latter. This is exactly the view I suggested above: that a threat to kill boosts the amount of force permitted in self-defense, but only somewhat. And it applies naturally to the case of aggression, where the threat to kill citizens who resist an assault on their political rights can likewise boost the amount of force they may use in response.

Finally, aggression threatens more than only citizens' right of political self-determination. To see how, consider the law of individual self-defense. Most jurisdictions limit the right of self-defense by proportionality conditions, and these

²⁹ McMahan, "Innocence, Self-Defense and Killing in War," p. 196.

³⁰ Norman, *Ethics, Killing and War*, pp. 130–31; Rodin, *War and Self-Defense*, pp. 132–38.

usually allow less force in defense of one's property than in defense of one's person. Whereas a victim may kill to prevent himself from being killed or badly injured, he may not kill to protect his property. But the law often makes a partial exception of a person's home. On the ground that "a man's house is his castle," it allows more force to be used in protecting one's home than in protecting other forms of property, in some jurisdictions even allowing killing to prevent mere forcible entry into one's home.³¹ It also allows a person to stand and resist an attack in his home that he would be required to retreat from elsewhere. This view has an intuitive rationale. If rape is a serious crime, it is not only because of the bodily harm it causes; it is also and more importantly because it violates what should be most intimate and private to a person. Similarly, though to a lesser degree, forcible entry into a person's home invades space that should be personal; this is why its victims feel violated. While the parallel with rape must not be exaggerated, it seems that just as in that case the violation of intimate space increases the seriousness of the crime and the amount of force permitted to prevent it, so invasion of one's home justifies more defensive force than other crimes against property.

A similar idea applies to international aggression. As the recent literature on nationalism has underscored, a nation's citizens typically regard it as a kind of home. They feel emotionally attached to its landscape, architecture, and cultural life, some of which are threatened by aggression. They also feel attached to its political institutions, seeing them as another aspect of their national home and resenting interferences with them. The culture of English-speaking Canada does not differ radically from that of the United States, yet most Canadians would find the armed incorporation of Canada into the United States and the replacement of Canadian political institutions by those of the United States a violation similar in kind to intrusion by a burglar into their home. For them, a U.S. attack would be an invasion not only literally but also metaphorically. But then it is a mistake to see the only rights of citizens threatened by aggression as rights of political self-determination; they also include the right to be secure in a political and cultural home. Just as in individual defense the protection of a home justifies some additional force, so protecting a political home does so in the morality of war.

Three features, then, differentiate aggression against a nation from interference with one person's right to vote: the large number of people whose rights are threatened, the fact that aggression is backed by a conditional threat to kill, and the fact that the attack is not only on rights of political participation but also invades a national home. Even if no one of these features by itself justifies killing, together they surely do. Given everything that aggression threatens, killing to prevent it need not be disproportionate.

³¹ See Stuart P. Green, "Castles and Carjackers: Proportionality and the Use of Deadly Force in Defense of Dwellings and Vehicles," *University of Illinois Law Review* 1999 (1999): 1-41. In endorsing the general idea that more defensive force is permitted in one's home I do not endorse these extreme "make my day" laws.

One would like to go further and say more precisely when defending sovereignty justifies war. If the prospect of global nuclear war makes resisting aggression wrong, are there less catastrophic effects that do the same? And what if the benefits of resisting are not so great? The nation attacked can be small, like Kuwait, so not many citizens' rights are threatened; undemocratic, again like Kuwait, so self-determination is not at issue; or not one whose citizens feel emotionally attached to its institutions. Do these factors reduce the justification for war? Or what if the aggressor seeks not to supplant the entire government but only to occupy a small, sparsely inhabited territory like the Falklands? In many of these cases the contributing cause of deterrence may favor fighting; the best way to prevent future, more serious aggression may be to resist even comparatively trivial aggressions now. But one would like to know what is intrinsically proportionate in these cases, and that is difficult to decide. Philosophy is most helpful in weighing competing moral considerations when it can find some more abstract value that underlies them and see how far each instantiates that value. But the considerations in play here seem irreducibly diverse: political self-determination and the protection of a national home on the one side, death and suffering on the other. This leaves their comparison to direct intuition, and, although that yields definite results in some cases, such as Czechoslovakia in 1968 or the invasion of a large democracy, it does not do so in the intermediate cases described above. It is therefore difficult to pin down more exactly how much force the defense of sovereignty allows. Nonetheless, I hope to have vindicated the commonsense view that at least sometimes, and certainly when a large, popular democracy faces total political absorption, national self-defense can justify lethal military force.

V. Weighing Goods and Evils: Comparing Lives

My second issue concerns how a nation should weigh lives when it kills some enemy citizens in order to save citizens of its own. This can be an *ad bellum* issue when a just cause for war is to prevent terrorist attacks like those of September 11, 2001; it also arises in the *jus in bello*, where soldiers must often choose between tactics that will cause more or fewer enemy casualties at the cost of more or fewer casualties for themselves. There is obviously no precise formula for making these choices, such as that 2.7 enemy lives equal one of ours. But we can try to describe the general parameters within which they should be made. Since there are two main categories of personnel on each side, combatants and noncombatants, there are four relevant comparisons: our soldiers against their soldiers, our civilians against their soldiers, our civilians against their civilians, and our soldiers against their civilians.³²

³² Further relevant categories are those of allied soldiers and civilians, and neutral soldiers and civilians. I leave these aside in this article.

In bello proportionality as standardly understood seems to allow a nation to kill virtually any number of enemy soldiers to save just one of its own soldiers. Once a war has begun, enemy soldiers are essentially free targets that one may attack at any time.³³ The *in bello* necessity condition forbids killing them wantonly, or for no military purpose. But if killing enemy soldiers now will prevent them from killing one of our soldiers in the future, it seems we may kill almost any number to achieve that end. Radical though it seems, this claim mirrors one from the morality of self-defense, where a person may kill any number of attackers if that is necessary to save his own or another's life. It also fits the charges of disproportionality leveled against the Gulf War, which concerned only the number of Iraqi civilians killed and not the number of Iraqi soldiers.³⁴ Or consider the movie *Saving Private Ryan*, in which a troop of U.S. soldiers rescue a fellow soldier caught behind enemy lines. There is no suggestion in the movie or in the common response to it that there is some number of German soldiers such that the troop must be careful not to kill more than that number while saving Ryan. Some theorists have argued that the traditional distinction between combatants and noncombatants should be rejected, on the ground that morally innocent conscript soldiers are less legitimate targets of force than civilians who culpably contributed to the start of an unjust war.³⁵ Although important, this argument raises issues beyond the scope of this chapter; and if we assume the traditional distinction or consider only volunteer enemy soldiers, it seems we may kill virtually any number to save one of our soldiers.³⁶

If so, however, we may also kill virtually any number of enemy soldiers to save one of our civilians. A government's duty to protect its civilians is surely as great as its duty to protect its soldiers; any preference it shows the latter it may also show the former. In addition, civilians have not, by volunteering for military service, accepted any risk of dying in war, so the government's responsibility to them is if anything greater. There may be limits on the priority a government may give its civilians' lives in the *jus ad bellum*. While it is permissible to initiate a war that will kill enemy soldiers to prevent large-scale attacks on our civilians, this may not be permissible to save one or two civilians. Once war has begun, however, it seems the priority is virtually absolute. If we can prevent an attack that will kill one of our civilians by killing a number of enemy soldiers, it seems we may do so almost whatever that number is.

³³ See Walzer, *Just and Unjust Wars*, pp. 138–51.

³⁴ After the war Gen. Norman Schwarzkopf expressed contempt for Saddam Hussein's willingness to squander Iraqi soldiers' lives. But he thought the responsibility to protect those soldiers belonged only to Saddam; Schwarzkopf's own duty was only to U.S. soldiers. See Jack Anderson and Dale Van Atta, *Stormin' Norman: An American Hero* (New York: Zebra Books, 1991), p. 164, cited in A. J. Coates, *The Ethics of War* (Manchester: Manchester University Press, 1997), p. 221.

³⁵ McMahan, "Innocence, Self-Defense and Killing in War," *passim*; Mellow, "A Critique of Just War Theory," *passim*.

³⁶ I say "virtually" because there may be some number (1,000? 10,000?) such that it is wrong to kill that many enemy soldiers to save one of ours. My point is simply that even if there is such a number, it is very high.

The more difficult issues concern weighing enemy civilian lives, first against our civilians' lives. Whereas enemy soldiers have, in becoming soldiers, given up certain protections and become legitimate targets of force, enemy civilians retain those protections, and one conclusion is that their lives must be weighed equally against those of our civilians. In 2001 many watched the death toll of Afghan civilians with the hope that it would not exceed the 3,000 Americans killed on September 11; similarly, some critics condemn Israel's attacks on suicide bombers in the occupied territories for killing more Palestinian civilians than the bombers have killed Israelis. Both these claims get one term of the moral comparison wrong. In the Afghan case the relevant U.S. number is not that of civilians killed on September 11; their lives were already lost. It is the number of U.S. civilians saved by the war, or the number of additional lives that would have been lost to terrorism had the war not been fought. In the Israeli case it is likewise the number of additional terrorist victims there would have been without the counterattacks. Setting this aside, however, both views assume that civilian lives on the two sides must be weighed equally.

The idea of equal weighting is familiar from moral views such as utilitarianism, which require equal consideration of all people's interests. But these views are sharply at odds with commonsense morality, which does not tell a father to care no more about his daughter than about a stranger. On the contrary, it says he may and should give his daughter's welfare considerably greater weight, so if he has a choice between saving his daughter's life and those of several strangers, he may and should do the former.³⁷ The relations among citizens of a nation are not as close as between parents and children, and the partiality they justify is not as strong. But common sense still calls for some partiality toward fellow citizens and certainly demands that partiality of governments. Although they have some duty to relieve poverty in other countries, they have a stronger duty to do so in their own; in formulating trade, immigration, and other policies governments should consider primarily the effects on their citizens.³⁸ Applied to just war theory, this view says a government should weigh its own civilians' lives more heavily than enemy civilians', and may therefore kill more of the latter if that is necessary to save somewhat fewer of the former.

It may be objected that this view mistakes the proper limits of justified partiality. Common sense permits us to prefer those closer to us when we are giving benefits, but not when the issue is causing harm. On the contrary, it gives everyone equal rights against such harm and requires those rights to be equally respected.

³⁷ Commonsense morality therefore incorporates what C. D. Broad called "self-referential altruism"; see his "Self and Others," in *Broad's Critical Essays in Moral Philosophy*, ed. David Cheney (London: George Allen & Unwin, 1971), pp. 262–82.

³⁸ I defend a moderate degree of national partiality in "The Justification of National Partiality," in Robert McKim and Jeff McMahan, eds., *The Morality of Nationalism* (New York: Oxford University Press, 1997), pp. 139–57.

A father may and even should prefer saving his daughter's life to saving five strangers, but he may not kill those strangers in order to save his daughter. The same goes for governments. They may prefer their citizens' interests when giving benefits like those of poverty relief or trade policy, but not when killing.

Despite its appeal to equality, this argument would have very restrictive implications in the context of war. After all, a father is not only forbidden to kill five strangers to save his daughter; he is also forbidden to kill one stranger to save five daughters. To adapt a familiar example, if he has five daughters who need different organ transplants and no organs are available, he is not permitted to kill one innocent person in order to divide up her organs among his daughters. Applied to war, then, the argument would make it disproportionate to kill a much smaller number of enemy civilians in the course of saving a much larger number of one's own. This is counterintuitive, and the reason is that it mistakes the important distinction in just war theory. This is not the distinction between causing harm and failing to prevent it; it is the distinction between targeting people for harm and harming them collaterally, which is a distinction within the category of causing. This latter distinction is usually understood using the doctrine of double effect, so it becomes the distinction between intending harm as an end or means and merely foreseeing that harm will result. But whatever its exact basis (and others have been proposed),³⁹ the targeted/collateral distinction is central to just war theory, so to test our view about weighing civilians' lives we need a nonmilitary example that involves it. Imagine that a victim is being attacked by an aggressor and that the only way a third party can save the victim's life is by throwing a grenade that will kill the attacker and also, unavoidably, an innocent bystander. It is arguable that if the third party is unrelated to any of the other participants he may not throw the grenade, and in particular may not prefer the victim's innocent life to the bystander's. But now imagine that the defender is the victim's father. It seems to me that he may throw the grenade, and may do so even if this will kill some number of bystanders greater than one. If he is not aiming at the bystanders but killing them collaterally, he may show some preference for his daughter. This claim will be contested by some. Thus, Judith Jarvis Thomson has denied that one may kill a bystander while defending not only a loved one but also oneself.⁴⁰ But she may feel forced to this conclusion by her rejection of the double effect distinction, and if we accept either that distinction or some other between targeted and

³⁹ F. M. Kamm has proposed grounding the distinction in a more complex causal condition; see her "Justifications for Killing Noncombatants in War," *Midwest Studies in Philosophy* 24 (2000): 219–28. But her condition implies that if we drop a bomb on a factory and a piece of the bomb flies through the air and kills a civilian, this is forbidden; whereas if we drop a bomb and a piece of the factory flies through the air and kills a civilian, that is not. I take it this is absurd. In my view double effect gives the best grounding for the targeted/collateral distinction, but I cannot rule out a priori the possibility of alternative groundings.

⁴⁰ Judith Jarvis Thomson, "Self-Defense," *Philosophy & Public Affairs* 20 (1991): 289–90, 296, 307–08.

collateral harms, as both just war theory and international law do, we can use it to justify the more permissive claim that the father may kill some bystanders if that is unavoidable in saving his daughter. And I think most will find this claim more intuitive than Thomson's; surely few in the father's position would fail to save their daughter. The fact that he is killing rather than failing to save is not irrelevant; it still plays a significant moral role, and in particular reduces the degree of partiality he may show below what would be permitted if he were merely distributing benefits. To put it a little technically, there is some number such that he would be permitted to save his daughter rather than save that number of strangers but may not kill that number of strangers as a side effect of saving his daughter. Given that limit, however, he is permitted to show some partiality toward his daughter even when he does kill bystanders.

The situation of a nation weighing its own against enemy civilians' lives is analogous. The nation is, say, attacking a government that has sponsored terrorist attacks against its citizens and finds that, while directing force only at that government's agents, it will unavoidably kill some enemy civilians. I think that in this case the nation's government is permitted to give somewhat greater weight to its own civilians' lives, and the case for partiality here may even be stronger than in that of individual defense. Even Thomson, who rejects the targeted/collateral distinction in individual cases, acknowledges that it has considerable intuitive force in the context of war, force that she needs to but cannot explain away.⁴¹ In fact the distinction is implicit in the very idea of *in bello* proportionality. The point of the *in bello* condition, recall, is to determine when acts that are not forbidden in themselves—for example, just killings of noncombatants—are forbidden because they cause excessive harm. But this issue would not arise unless some acts of killing were *not* forbidden in themselves, that is, unless there was something like a targeted/collateral distinction. So any discussion of *in bello* proportionality must assume the distinction, and once it is in place there is room for a government to give somewhat greater weight to its own civilians' lives. The fact that the government's acts will kill enemy civilians remains important, and in particular permits less partiality than is appropriate in trade or immigration policy. I wish I could say more precisely what this degree is, or how many enemy civilian deaths are proportionate side effects of saving one of our civilians. But I do think that when weighing its own civilians' lives against those of enemy civilians it will merely collaterally kill, a nation may give some preference to the former: more than zero preference, but not as much as is permitted when no killing is involved. If the nation is trying to prevent terrorist attacks like those of September 11, 2001, then tragic though the result will be, and assuming the nation makes serious efforts to minimize collateral harm, it may kill somewhat more enemy civilians if that is unavoidable in saving a smaller number of its own.

⁴¹ Thomson, "Self-Defense," pp. 292–98.

The final case of weighing, of our soldiers against their civilians, may have the most practical importance. Since Vietnam, U.S. military tactics have had as a central aim keeping U.S. casualties to a minimum. In the Gulf War this led to the massive bombing of Iraq's infrastructure that preceded the ground war; in Kosovo it caused NATO to fly its planes above 15,000 feet, where they were beyond the reach of Yugoslav anti-aircraft fire but where their bombing was inevitably less accurate. Many critics charge that this desire to minimize U.S. military casualties has led to excessive civilian casualties on the other side. To assess this complaint we must know what an acceptable tradeoff between our military and enemy civilian lives would be.

I believe a nation may again give some extra weight to its soldiers' lives because they are its own. Like its civilians, its soldiers are citizens and may be given some preference on that basis. In addition, military commanders typically do and should feel special responsibility for the troops under their command. This is not just a matter of preferring their lives to enemy soldiers', but extends to protecting them from natural dangers, friendly fire, and, perhaps most important, having their lives wasted in ill-planned campaigns. It is no exaggeration to say that for a responsible military commander his troops have something like the status of family. But in this final weighing there is a competing consideration. Although our soldiers are ours, they are also soldiers, which means they are legitimate targets of military force and their deaths are an expected consequence of war as civilians' deaths are not. Paul Christopher has emphasized this point, saying that "risking one's life is part of what it means to be a soldier," and concluding that our soldiers' lives weigh less in the moral balance than do enemy civilians'.⁴² His argument is especially telling in a nation with a volunteer military, since its soldiers chose military service and so voluntarily accepted a risk of death as enemy civilians did not. This is an especially clear reason to prefer the lives of enemy civilians to those of our soldiers: while they are not ours and therefore count for less, they did not choose to become soldiers and therefore count for more. To resolve our final issue about weighing we must balance these competing considerations against each other.

This would again be easiest if we could find some more abstract value underlying the two, but they seem too diverse. One involves the special relationship of conationality; the other, at its most compelling, the voluntary acceptance of risk. So there seems no alternative to a direct intuitive assessment of their weights, and this is again difficult to make. I do not see either consideration as clearly tipping the scales in favor of one category of lives. I therefore feel forced to treat them as of approximately equal weight, so our soldiers' and enemy civilians' lives count roughly equally. While a nation may prefer its own civilians' lives to those of enemy civilians, it may not do the same with its soldiers' lives. Instead, it must trade those

⁴² Christopher, *The Ethics of War and Peace*, p. 165; see also p. 95. Christopher rejects the idea of national partiality, holding that enemy civilians' lives must be weighed equally against our civilians' (pp. 95, 97, 165–66, 172–73).

off against enemy civilians' lives at roughly one to one. This is not to say that an act that kills 101 civilians as a side effect of saving 100 soldiers is necessarily disproportionate; the comparisons cannot be that precise. But it does imply that any act that kills significantly more civilians than it saves soldiers is morally impermissible.

This view does not yield as restrictive a version of *in bello* proportionality as some may wish, but it still raises serious moral questions about recent military campaigns. In assessing the Gulf War bombing, we may have to discount the resulting deaths for Saddam's failure to repair his country's infrastructure; we must also compare these deaths not with the actual number of coalition casualties but with the additional casualties there would have been without the bombing, and in making a subjective assessment we must consider how matters appeared to coalition commanders at the time, when the expected casualties from the ground war were greater than actually resulted. Even so, it is difficult to see the bombing campaign as proportionate. In particular, it is hard to believe that a reduction in the bombing resulting in, say, 30 percent fewer civilian deaths would have had much effect on coalition casualties. A report to the International Criminal Tribunal for the Former Yugoslavia, investigating whether NATO should be prosecuted for its conduct of the Kosovo war, concluded that flying above 15,000 feet is not a war crime. Given the vagueness of the legal definitions of proportionality, that is probably true, but there remain serious moral concerns. In particular, it is hard to believe that occasionally flying below 15,000 feet, especially when identifying targets by the naked eye, would have cost more NATO military lives than it saved Yugoslav civilians.⁴³ Again, even a one-to-one tradeoff between our military and enemy civilian lives raises questions about the conduct of this war.

For individual government officials in these cases there may be a mitigating factor. If U.S. leaders have tried to minimize U.S. military casualties, it is partly for fear that otherwise U.S. public opinion will oppose war and make it harder to fight. Henry Kissinger urged this type of point in response to proposals that the Gulf War be delayed to give sanctions more time to work; if one waited, he said, "a credible military option probably would no longer exist," because public support would have eroded to the point where war was no longer politically possible.⁴⁴ When politicians say a particular way of fighting is "politically impossible" they sometimes mean only that it will reduce their popularity and chance of reelection. But at other times what they say is literally true: too many casualties for their side will spark public and political opposition to the war that makes it impossible to win. In that

⁴³ In a much-publicized incident, NATO aircraft bombed a convoy of Albanian refugees, killing 70 to 75 of them, in the mistaken belief that they were Serb military forces. The pilots were flying at 15,000 feet and viewing their target with the naked eye; see International Criminal Tribunal for the Former Yugoslavia, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia*, sec. 63–70, available at <http://www.un.org/icty/pressreal/natoo61300.htm>.

⁴⁴ Henry Kissinger, testimony to U.S. Senate Armed Services Committee, quoted in Coates, *The Ethics of War*, p. 197.

case what weighs against enemy civilians' lives in the politicians' deliberations is not just their own soldiers' lives but any chance of achieving the war's just causes—that is, the whole positive point of the war. If winning the war with less collateral damage is not possible, the politicians' choice may actually be proportionate, achieving the goods of war at the least possible cost.⁴⁵ But the same is not true of the nation as a whole, comprising the government and the public for whom it acts. Since a nation is not constrained by others' reactions, when it kills more enemy civilians to save a smaller number of its soldiers it acts on suspect nationalist preferences and violates proportionality.

VI. Conclusion

Judgments about proportionality in war cannot be made simply or mechanically. Even if we have specified the types of good and evil relevant to these judgments, we must identify the specific effects of a given war or act, which requires comparing that war or act with alternatives that are merely hypothetical and can only be estimated given our available evidence. In addition, once the relevant effects are identified they must be weighed against each other, which often involves some indeterminacy because they are of irreducibly different types. But that proportionality judgments involve some indeterminacy does not mean they can never be made. The common argument that what cannot be measured precisely cannot be measured at all is as fallacious here as elsewhere. And we surely can make some determinate proportionality judgments. Thus, we can say that a conventional war fought to defend a nation's sovereignty against aggression is normally proportional, while a tactic that kills many enemy civilians rather than sacrifice a few soldiers is not. In the first case the relevant goods clearly outweigh the relevant evils; in the second they do not.

Postscript

This chapter claims confidently that the proportionality requirement in the international law of war allows no discounting of harms for wrongful contributing acts

⁴⁵ There is an issue about intervening causes here. If a state's leaders spare enemy civilians and therefore lose the war, that result follows only because of a response by their public that is morally wrong, because it involves excessive partiality to their own soldiers. If there is discounting for others' wrongful choices, the leaders are not fully responsible for the loss of the war and may have to accept it rather than kill the civilians; if all resulting evils count, the leaders may be justified in doing what is necessary to win. This is ironic, since in our earlier discussion the view that discounts for intervening causes had more hawkish implications, for example, allowing more collateral damage from bombing than if one counted all actual harms. Here the view that discounts has more dovish implications, requiring a state's leaders not to fight when the alternative may allow them to do so.

by one's enemy, as when they have placed a military target close to civilians. This is in fact a matter of controversy. While some legal analysts do take that view, others do not. An analysis of the laws of war published by the U.K. Ministry of Defence says that although proportionality still applies when defenders wrongfully place military objectives near civilians, the latter "is a factor to be taken into account in favor of the attackers in considering the legality of attacks on those objectives."⁴⁶ There is also a serious error in argumentation near the end of the essay. I claim that, if competing considerations give our soldiers' and enemy civilians' lives roughly equal weight, we should weigh the two roughly equally when assessing *in bello* proportionality. But this does not follow. The chapter earlier argued that, although the doing/allowing distinction is not the most important in the morality of war, it still has some weight. That is why a nation may not give as much priority to its own civilians' over enemy civilians' interests in war, when it may collaterally kill the enemy civilians, as it may in its trade or immigration policies, which only deny noncitizens benefits. But then the same point should apply to tradeoffs between our soldiers and enemy civilians in war: the fact that we will actively kill civilians in the course of protecting our soldiers should give the former's lives somewhat greater weight, so the collateral killing of enemy civilians is justified only if it will save a somewhat greater number of our soldiers' lives. This conclusion seems both more intuitive than the one defended in the chapter and closer to standard thinking about the morality of war.

⁴⁶ *The Manual of the Law of Armed Conflict* (Oxford: Oxford University Press, 2004), sec. 2.7.2; see also 5.22.1.

{ REFERENCES }

- Ackerman, Bruce, *Social Justice in the Liberal State* (New Haven: Yale University Press, 1980).
- Adelman, Tom, *The Long Ball: The Summer of 75—Spaceman, Catfish, Charlie Hustle, and the Greatest World Series Ever Played* (New York: Back Bay Books, 2003), p. 313.
- Anderson, Jack, and Dale Van Atta, *Stormin' Norman: An American Hero* (New York: Zebra Books, 1991), p. 164.
- Annas, Julia, "Virtue Ethics," in *Oxford Handbook of Ethical Theory*, ed. D. Copp (Oxford: Clarendon Press, 2005), pp. 515–36.
- Anscombe, G. E. M., "Modern Moral Philosophy," *Philosophy* 33 (1958): 1–19.
- Aquinas, St. Thomas, *Summa Theologica*, trans. Fathers of the English Dominican Province, 4 vols. (Westminster, Md.: Christian Classics, 1981).
- Arpal, Nomy, *Unprincipled Virtue: An Inquiry into Moral Agency* (New York: Oxford University Press, 2003).
- Aristotle, *Nicomachean Ethics*, trans. W. D. Ross and J. O. Urmson, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984).
- , *Eudemian Ethics*, trans. J. Solomon, in *The Complete Works of Aristotle*, 2 vols., ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984).
- Arrhenius, Gustaf, "An Impossibility Theorem for Welfarist Axiologies," *Economics and Philosophy* 16 (2000): 247–66.
- Audi, Robert, "Intrinsic Value and Moral Obligation," *The Southern Journal of Philosophy* 35 (1997): 140.
- Ayer, J., *Language, Truth, and Logic* (London: Gollancz, 1936).
- Barry, Brian, *Political Argument* (London: Routledge & Kegan Paul, 1965).
- Blackburn, Simon, *Ruling Passions; A Theory of Practical Reasoning* (Oxford: Clarendon Press, 1998).
- Blackorby, Charles, Walter Bossert, and David Donaldson, "Critical-Level Utilitarianism and the Population-Ethics Dilemma," *Economics and Philosophy* 13 (1997): 197–230.
- Blackstone, Sir William, *Commentaries on the Laws of England* (Chicago: University of Chicago Press, 2002).
- Bradford, Gwendolyn, "Kudos for Ludus: Games and Value Theory," *Noesis* 6 (2003): 15–28.
- Bradley, F. H., *The Principles of Logic*, 2nd ed. (Oxford: Oxford University Press, 1922).
- , "Mr. Sidgwick's Hedonism," in *Collected Essays*, vol. 1 (New York: Oxford University Press, 1935).
- Brennan, Samantha, "Thresholds for Rights," *Southern Journal of Philosophy* 33 (1995): 143–68.
- Brentano, Franz, *The Origin of Our Knowledge of Right and Wrong*, trans. Roderick M. Chisholm and Elizabeth Schneewind (New York: Routledge, 1969).
- Brink, David O., "Self-Love and Altruism," *Social Philosophy and Policy* 14/1 (Winter 1997): 122–57.
- Broad, D., *Critical Essays in Moral Philosophy* (London: George Allen & Unwin, 1971).
- Carritt, E. F., *Ethical and Political Thinking* (Oxford: Clarendon Press, 1947).
- Chisholm, R. M., *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986).
- Christopher, Paul, *The Ethics of War and Peace: An Introduction to Legal and Moral Issues*, 2d ed. (Upper Saddle River, N.J.: Prentice-Hall, 1999).
- Coates, A. J., *The Ethics of War* (Manchester: Manchester University Press, 1997).

- Crisp, R., and M. Slote, "Introduction," in *Virtue Ethics*, ed. R. Crisp and M. Slote (Oxford: Oxford University Press, 1997).
- Dancy, Jonathan, *Moral Reasons* (Oxford: Blackwell, 1993).
- Darwin, Major Leonard, "Memories of Down House," in *The Nineteenth Century*, 106 (1929): 119–20, quoted in William Irvine, *Apes, Angels, and Victorians* (New York: Time, 1963).
- De Sousa, Ronald, "The Good and the True," *Mind* 83 (1974): 534–51.
- , *The Rationality of Emotion* (Cambridge, Mass.: MIT Press, 1990).
- Dolinko, D., "How to Criticize the Death Penalty," *Journal of Criminal Law and Criminology* 77 (1986): 546–601.
- , "Justice in the Age of Sentencing Guidelines," *Ethics* 110 (2000): 563–85.
- Doris, John, *Lack of Character: Personality and Moral Behavior* (Cambridge: Cambridge University Press, 2002).
- Dworkin, Gerald, "Is More Choice Better than Less?" in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds., *Midwest Studies in Philosophy* 7 (1982): 47–61.
- Dworkin, Ronald, "Liberalism," in *Public and Private Morality*, ed. Stuart Hampshire (Cambridge: Cambridge University Press, 1977).
- , "Foundations of Liberal Equality," *The Tanner Lectures on Human Values*, vol. 11 (Salt Lake City: University of Utah Press, 1990).
- Ewing, A. C., *The Morality of Punishment* (London: Kegan Paul, Trench, Tubner & Co.: 1929).
- Feinberg, Joel, *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, N.J.: Princeton University Press, 1970).
- , *Social Philosophy* (Englewood Cliffs, NJ: Prentice-Hall, 1973).
- , *Rights, Justice, and the Bounds of Liberty* (Princeton, N.J.: Princeton University Press, 1980).
- Feldman, Fred, *Utilitarianism, Hedonism, and Desert* (Cambridge: Cambridge University Press, 1997).
- Fletcher, George P., *Rethinking Criminal Law* (Boston: Little, Brown, 1978).
- Foot, Philippa, "Moral Beliefs," *Proceedings of the Aristotelian Society* 59 (1958–9): 83–104.
- , *Virtues and Vices* (Berkeley, CA: University of California Press, 1973).
- Frank, R. H., *Choosing the Right Pond: Human Behavior and the Quest for Status* (New York: Oxford University Press, 1985).
- Frankena, William K., *Ethics*, 2nd ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1973).
- Freedman, Lawrence, and Efraim Karsh, *The Gulf War 1990–91: Diplomacy and War in the New World Order* (London: Faber & Faber, 1993).
- Furman v. Georgia*, 408 US 239 (1972), quoted in H. A. Bedau, ed., *The Death Penalty in America*, 3rd ed. (New York: Oxford University Press, 1982).
- Gardam, Judith Gail, "Proportionality and Force in International Law," *American Journal of International Law* 87 (1993): 392–94.
- Gardner, Howard, *Frames of Mind: The Theory of Multiple Intelligences* (New York: Basic Books, 1983).
- Garcia, J. L. A., "Goods and Evils," *Philosophy and Phenomenological Research*, 47 (1987): 385–412.
- , "The Primacy of the Virtuous," *Philosophia*, 20 (1990): 69–91.
- Gauthier, David, *Morals by Agreement* (Oxford: Clarendon Press, 1986).
- Gee, Marcus, "Stick to your Poetry and Pianos," *Globe and Mail*, Sept. 27, 2002, p. A17.
- Gewirth, Alan, *Reason and Morality* (Chicago: University of Chicago Press, 1978).
- Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Mass.: Harvard University Press, 1990).
- Globe and Mail*, "Marines Wade into Dreaded Urban Battle," March 25, 2003.
- Glover, Jonathan, *Causing Death and Saving Lives* (Harmondsworth: Penguin Books, 1977).
- Goldman, Alan H., "The Paradox of Punishment," *Philosophy and Public Affairs* 9 (1979–80): 42–58.
- Green, Stuart P., "Castles and Carjackers: Proportionality and the Use of Deadly Force in Defense of Dwellings and Vehicles," *University of Illinois Law Review* 1999 (1999): 1–41.
- Griffin, James, "Is Unhappiness Morally More Important than Happiness?" *Philosophical Quarterly* 29 (1979): 47–55.

- , *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Clarendon Press, 1986).
- Hamilton, William, *The Metaphysics of Sir William Hamilton*, ed. F. Bowen (Boston: John Allyn, 1872).
- Hare, R. M., *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981).
- Harman, Gilbert, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99 (1999): 315–31.
- Harman, G., and J. J. Thomson, *Moral Relativism and Moral Objectivity* (Oxford: Blackwell, 1996).
- Hart, H. L. A., "Are There Any Natural Rights?" *Philosophical Review* 64 (1955): 175–191.
- , *Punishment and Responsibility* (Oxford: Clarendon Press, 1968).
- Hoffman, J., "A New Theory of Comparative and Noncomparative Justice," *Philosophical Studies* 70 (1993): 165–83.
- Hooker, Brad, *Ideal Code/Real World* (Oxford: Clarendon Press, 2000).
- Hurka, Thomas, "Virtue as Loving the Good," *Social Philosophy and Policy* 9 (1992): 149–68.
- , *Perfectionism* (New York: Oxford University Press, 1993).
- , "The Justification of National Partiality," in *The Morality of Nationalism*, ed. Robert McKim and Jeff McMahan (New York: Oxford University Press, 1997).
- , "Self-Interest, Altruism, and Virtue," *Social Philosophy and Policy* 14 (1997): 286–307.
- , "How Great a Good Is Virtue?" *The Journal of Philosophy* 95 (1998): 181–203.
- , "Two Kinds of Organic Unity," *Journal of Ethics* 2 (1998): 299–320.
- , *Virtue, Vice, and Value* (New York: Oxford University Press, 2001).
- Hurley, Susan, *Natural Reasons* (New York: Oxford University Press, 1989).
- Hursthouse, Rosalind, "Virtue Theory and Abortion," *Philosophy and Public Affairs* 20 (1991): 223–46.
- , *On Virtue Ethics* (Oxford: Clarendon Press, 1999).
- International Criminal Tribunal for the Former Yugoslavia, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia*, sec. 63–70, available at <http://www.un.org/icty/pressreal/natoo61300.htm>.
- Johnson, James Turner, *Just War Tradition and the Restraint of War* (Princeton, N.J.: Princeton University Press, 1981).
- , *Can Modern War Be Just?* (New Haven: Yale University Press, 1984).
- , *Morality and Contemporary Warfare* (New Haven: Yale University Press, 1999).
- Kagan, Shelly, "The Additive Fallacy," *Ethics* 99 (1988): 5–31.
- , *The Limits of Morality* (Oxford: Clarendon Press, 1989).
- , "Replies to My Critics," *Philosophy and Phenomenological Research* 51 (1991): 919–20.
- , "The Limits of Well-Being," *Social Philosophy and Policy* 9 (1992): 169–89.
- , "Me and My Life," *Proceedings of the Aristotelian Society* 94 (1993/94): 318–23.
- , "Rethinking Intrinsic Value," *Journal of Ethics* 2 (1998): 277–97.
- , "Equality and Desert," in *What Do We Deserve? Readings on Justice and Desert*, ed. Louis P. Pojman and Owen McLeod (New York: Oxford University Press, 1999).
- , "The Grasshopper, Aristotle, Bob Adams, and Me" (unpublished ms.).
- Kamm, F. M., *Morality/Mortality, vol. 1: Death and Whom to Save from It* (New York: Oxford University Press, 1993).
- , *Morality/Mortality, vol. 2: Right, Duties, and Status* (New York: Oxford University Press, 1996).
- , "Justifications for Killing Noncombatants in War," *Midwest Studies in Philosophy* 24 (2000): 219–28.
- Kant, Immanuel, *The Metaphysical Elements of Justice: Part I of the Metaphysics of Morals*, trans. John Ladd (Indianapolis: The Liberal Arts Press, 1965).
- , *Allgemeine Naturgeschichte und Theorie des Himmels* (1755), in *Kants gesammelte Schriften*, 24 vols. (Berlin: Königlich Preussischen Akademie der Wissenschaften, 1902–66).
- Kavka, Gregory S., "The Paradox of Future Individuals," *Philosophy and Public Affairs* 11 (1982): 93–112.

- Kekes, John, *The Morality of Pluralism* (Princeton, N.J.: Princeton University Press, 1993).
- Korsgaard, Christine M., "Two Distinctions in Goodness," *Philosophical Review* 92 (1983): 169–95.
- , *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).
- Kymlicka, Will, *Contemporary Moral Philosophy* (Oxford: Clarendon Press, 1990).
- Lackey, Douglas P., *The Ethics of War and Peace* (Englewood Cliffs, N.J.: Prentice-Hall, 1989).
- Lamont, Julian, "Incentive Income, Deserved Income, and Economic Rents," *Journal of Political Philosophy* 5 (1997): 26–46.
- Locke, John, *Two Treatises of Government*, 2nd ed., ed. Peter Laslett (New York: Cambridge University Press, 1967).
- Lovejoy, A. O., *The Great Chain of Being* (Cambridge, Mass.: Harvard University Press, 1936).
- Lemos, Noah M., "High-minded Egoism and the Problem of Priggishness," *Mind* 93 (1984): 542–58.
- , *Intrinsic Value: Concept and Warrant* (Cambridge: Cambridge University Press, 1994).
- Luban, David, "Just War and Human Rights," *Philosophy & Public Affairs* 9 (1979/80): 160–81.
- MacIntyre, Alasdair, "Is Patriotism a Virtue?" Lindley Lecture (Lawrence, KS: University of Kansas, 1984).
- Margalit, Avishai, and Joseph Raz, "National Self-Determination," *Journal of Philosophy* 87 (1990): 439–61.
- Marshall, G., et al., "What Is and What Ought to Be: Popular Beliefs about Distributive Justice in Thirteen Countries," *European Sociological Review* 15 (1999): 349–67.
- Marx, Karl, *Grundrisse*, trans. M. Nicolaus (Harmondsworth: Penguin, 1973).
- , *Capital* (New York: Vintage, 1977).
- Marx, Karl, and Friedrich Engels, "The Communist Manifesto," in *Karl Marx: Selected Writings*, ed. D. McLellan (Oxford: Oxford University Press, 1977).
- , "The German Ideology," in *Karl Marx: Selected Writings*, ed. D. McLellan (New York: Oxford University Press, 1977).
- Mayerfeld, Jamie, "The Moral Asymmetry of Happiness and Suffering," *Southern Journal of Philosophy* 34 (1996): 317–38.
- , *Suffering and Moral Responsibility* (New York: Oxford University Press, 1999).
- McDowell, John, "Virtue and Reason," *The Monist* 62 (1979): 331–50.
- McKenna, Joseph C., "Ethics and War: A Catholic View," *American Political Science Review* 54 (1960): 647–58.
- McKerlie, Dennis, "Egalitarianism," *Dialogue* 23 (1984): 223–37.
- McLeod, Owen, "Desert and Wages," in *What Do We Deserve? Readings on Justice and Desert*, ed. Louis P. Pojman and Owen McLeod (New York: Oxford University Press, 1999).
- McMahan, Jeff, "Killing, Letting Die, and Withdrawing Aid," *Ethics* 103 (1993): 250–79.
- , "Innocence, Self-Defense and Killing in War," *Journal of Political Philosophy* 2 (1994): 193–221.
- , "The Limits of National Partiality," in *The Morality of Nationalism*, ed. Robert McKim and Jeff McMahan (New York: Oxford University Press, 1997): 107–38.
- McMahan, Jeff, and Robert McKim, "The Just War and the Gulf War," *Canadian Journal of Philosophy* 23 (1993): 501–41.
- McTaggart, J. M. E., *The Nature of Existence*, vol. 2 (Cambridge: Cambridge University Press, 1927).
- Mellow, David, "A Critique of Just War Theory," Ph.D. diss., University of Calgary, 2003.
- Mill, John Stuart, *Utilitarianism, Liberty, and Representative Government* (London: J. M. Dent, 1915).
- , *Utilitarianism*, ed. Oskar Piest (Indianapolis: Bobbs-Merrill, 1957).
- Miller, David, "The Ethical Significance of Nationality," *Ethics* 98 (1988): 649–50.
- , *Market, State and Community: Theoretical Foundations of Market Socialism* (Oxford: Clarendon Press, 1989).
- , *On Nationality* (Oxford: Clarendon Press, 1995).
- , *Principles of Social Justice* (Cambridge, Mass.: Harvard University Press, 1999).

- Montague, P. "Comparative and Non-Comparative Justice," *Philosophical Quarterly* 30 (1980): 131–40.
- Moore, G. E., *Principia Ethica* (Cambridge: Cambridge University Press, 1903).
- , "Mr. McTaggart's Ethics," *International Journal of Ethics* 13 (1903): 341–70.
- , *Philosophical Studies* (London: Routledge & Kegan Paul, 1922).
- , *Ethics* (Oxford: Oxford University Press, 1965).
- Morris, Herbert, "Persons and Punishment," *The Monist* 52 (1968): 475–501.
- Murphy, Liam B., *Moral Demands in Nonideal Theory* (New York: Oxford University Press, 2000).
- Nagel, Thomas, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970).
- , "Libertarianism without Foundations," *Yale Law Journal* 85 (1975): 136–49.
- , *Mortal Questions* (New York: Cambridge University Press, 1979).
- , *The View From Nowhere* (New York: Oxford University Press, 1986).
- Newman, John Henry Cardinal, *Certain Difficulties Felt by Anglicans in Catholic Teaching*, vol. 1 (London: Longmans, 1901).
- Nietzsche, Friedrich, *Beyond Good and Evil*, trans. Walter Kaufmann (New York: Vintage, 1966).
- , *The Will to Power*, trans. Walter Kaufmann and R. J. Hollingdale (New York: Vintage, 1968).
- , *On the Genealogy of Morals*, trans. W. Kaufmann and R. J. Hollingdale (New York: Vintage, 1969).
- , *The Gay Science*, trans. Walter Kaufmann (New York: Vintage, 1974).
- , *The Portable Nietzsche*, ed. W. Kaufmann (Harmondsworth: Penguin, 1976).
- Norman, Richard, *Ethics, Killing and War* (Cambridge: Cambridge University Press, 1995).
- Nozick, Robert, "Moral Complications and Moral Structures," *Natural Law Forum* 13 (1968): 1–50.
- , *Anarchy, State, and Utopia* (New York: Basic Books, 1974).
- , *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981).
- Nussbaum, Martha C., *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (Cambridge: Cambridge University Press, 1986).
- O'Brien, William V., *The Conduct of a Just and Limited War* (New York: Praeger, 1981).
- Parfit, Derek, *Reasons and Persons* (Oxford: Clarendon Press, 1984).
- , "On Giving Priority to the Worse-Off" (ms., All Souls College, Oxford, 1989).
- , "Equality or Priority?" The Lindley Lecture (Lawrence, KS: University of Kansas Press, 1995).
- Plato, *Republic*, trans. G. M. A. Grube (Indianapolis: Hackett, 1974).
- Pollack, Kenneth, *The Threatening Storm: The Case for Invading Iraq* (New York: Random House, 2002).
- Quinn, Warren S., "The Right to Threaten and the Right to Punish," *Philosophy and Public Affairs* 14 (1985): 327–73.
- , "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," *Philosophical Review* 98 (1989): 287–312.
- , "Actions, Intentions, and Consequences: The Doctrine of Double Effect," *Philosophy and Public Affairs* 18 (1989): 334–51.
- Rachels, James, "Active and Passive Euthanasia," *New England Journal of Medicine* 292 (1975): 78–80.
- Ramsey, Paul, *The Just War: Force and Political Responsibility* (New York: Scribner's, 1968).
- Rashdall, Hastings, *The Theory of Good and Evil*, 2 vols. (New York: Oxford University Press, 1907).
- Rawls, John, "Two Concepts of Rules," *Philosophical Review* 64 (1955): 3–32.
- , *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971).
- , "The Independence of Moral Theory," *Proceedings of the American Philosophical Association* 48 (1974–5): 5–22.
- Regan, Richard J., *Just War: Principles and Cases* (Washington, D.C.: Catholic University of America Press, 1996).
- Roberts, Adam, and Richard Guelff, eds., *Documents on the Laws of War*, 3d ed., (Oxford: Oxford University Press, 2000), p. 449.
- Rodin, David, *War and Self-Defense* (Oxford: Clarendon Press, 2002).

- Ross, W. D., *The Right and the Good* (Oxford: Clarendon Press, 1930).
- , *The Foundations of Ethics* (New York: Oxford University Press, 1939).
- , *Aristotle* (New York: Macmillan, 1949).
- Sadurski, W., *Giving Desert Its Due* (Dordrecht: Reidel, 1985).
- Sandel, Michael, "The Procedural Republic and the Unencumbered Self," *Political Theory* 12 (1984): 81–96.
- Scanlon, T. M., *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998).
- Scheffler, Samuel, *The Rejection of Consequentialism: A Philosophical Examination of the Considerations Underlying Rival Moral Conceptions* (Oxford: Clarendon Press, 1982).
- , "Justice and Desert in Liberal Theory," *California Law Review* 88 (2000): 965–90.
- , *Boundaries and Allegiances* (Oxford: Oxford University Press, 2001), 149–72.
- Sher, George, *Desert* (Princeton, N.J.: Princeton University Press, 1987).
- , *Beyond Neutrality: Perfectionism and Politics* (Cambridge: Cambridge University Press, 1997).
- Sidgwick, Henry, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907).
- Sikora, R. I., "Is It Wrong to Prevent the Existence of Future Generations?" in *Obligations to Future Generations*, ed. R. I. Sikora and Brian Barry (Philadelphia: Temple University Press, 1978).
- Singer, Peter, *Practical Ethics* (Cambridge: Cambridge University Press, 1979).
- Slote, Michael, *Common-Sense Morality and Consequentialism* (London: Routledge & Kegan Paul, 1985).
- , *Beyond Optimizing: A Study of Rational Choice* (Cambridge, Mass.: Harvard University Press, 1989).
- , "The Virtue in Self-interest," *Social Philosophy and Policy* 14 (1997): 273–274.
- Stevenson, Charles L., *Ethics and Language* (New Haven: Yale University Press, 1944).
- Stocker, Michael, *Plural and Conflicting Values* (Oxford: Clarendon Press, 1990).
- Suits, Bernard, *The Grasshopper: Games, Life and Utopia* (Toronto: University of Toronto Press, 1978).
- Swift, A., G. Marshall, and C. Burgoyne, "Which Road to Social Justice?" *Sociology Review* 1 (1992): 28–31.
- Tamir, Yael, *Liberal Nationalism* (Princeton: Princeton University Press, 1993).
- Taurek, John, "Should the Numbers Count?" *Philosophy and Public Affairs* 7 (1976–1977): 293–316.
- Taylor, Charles, *Multiculturalism and "The Politics of Recognition,"* ed. Amy Gutmann (Princeton: Princeton University Press, 1992).
- Temkin, Larry S., *Inequality* (New York: Oxford University Press, 1993).
- Thomson, Judith Jarvis, "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204–17.
- , "Self-Defense," *Philosophy & Public Affairs* 20 (1991): 283–310.
- , "The Right and the Good," *Journal of Philosophy* 94 (1997): 273–98.
- U.K. Ministry of Defence, *The Manual of the Law of Armed Conflict* (Oxford: Oxford University Press, 2004).
- U.S. Catholic Bishops, *The Challenge of Peace: God's Promise and Our Response*, reprinted in *Just War Theory*, ed. Jean Bethke Elshtain (New York: New York University Press, 1992).
- U.S. Department of the Army, *The Law of Land Warfare* (Field Manual 27–10, 1956).
- Van Den Haag, E., "In Defense of the Death Penalty: A Legal-Practical-Moral Analysis," *Criminal Law Bulletin* 14 (1978): 56.
- Volokh, Alexander, "n Guilty Men," *University of Pennsylvania Law Review* 146 (1997): 173–216.
- Von Humboldt, Wilhelm, *The Limits of State Action*, trans. J. W. Burrow (New York: Cambridge University Press, 1969).
- Walzer, Michael, *Spheres of Justice* (New York: Basic Books, 1983).
- , *Interpretation and Social Criticism* (Cambridge, Mass.: Harvard University Press, 1987).
- , *Just and Unjust Wars*, 2d ed. (New York: Basic Books, 1992).
- Warnock, G. J., *Contemporary Moral Philosophy* (London: Macmillan, 1967).

- , *The Object of Morality* (London: Methuen, 1971).
- Wente, Margaret, "Who's Killing the Children of Iraq?" *Globe and Mail*, Oct. 8, 2002, p. A23.
- Wiggins, David, "Weakness of Will, Commensurability, and the Objects of Deliberation and Desire," in *Essays on Aristotle's Ethics*, ed. A. O. Rorty (Berkeley: University of California Press, 1980).
- Watson, Gary, "On the Primacy of Character," in *Identity, Character, and Morality: Essays in Moral Psychology*, ed. O. Flanagan and A. O. Rorty (Cambridge, Mass.: MIT Press, 1990).
- Williams, Bernard, *Morality: An Introduction to Ethics* (New York: Harper & Row, 1972).
- , "A Critique of Utilitarianism," in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973).
- , *Problems of the Self* (Cambridge: Cambridge University Press, 1973).
- , "Utilitarianism and Moral Self-Indulgence," in Williams, *Moral Luck: Philosophical Papers 1973–80* (Cambridge: Cambridge University Press, 1981).
- Wittgenstein, Ludwig, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1972).
- Zimmerman, Michael J., "Virtual Intrinsic Value and the Principle of Organic Unities," *Philosophy and Phenomenological Research* 59 (1999): 653–66.

{ INDEX }

- achievement lines 45–52
achievement, value of 37–39, 60, 68–69, 74–75,
95, 100, 133–35, 185, 188–90
autonomy as an instance of 142–49
degrees of value in xi, 39, 147–48, 190–91
in games 191–98
aesthetic contemplation. *See* beauty, admiring
contemplation of
Afghanistan War 239, 244, 258
agency. *See* achievement, autonomy as
instance of
agent-relativity 5, 8, 11–13, 15–17, 20,
218, 224
Ali, Muhammad 34–35
Annas, Julia 180
antitheory 8–10
apathy. *See* sloth
Aquinas, St. Thomas 25–26, 34, 36, 37, 40
Aristotle xi, 8, 40–41, 95, 100, 129, 194–96,
198, 221
on knowledge, value of 37–38, 189
on virtue 55, 178, 181, 183, 193
art 37–38, 185, 188, 196–97, 227, 235
asymmetries in value 36, 113–35
deontological 130
desert/undesert 125–30
downshift basis of 120–23, 126–28
knowledge/false belief 133–34
limit 120, 124–25, 129
marginal-value basis of 115, 125–26
positive/negative goods 131–32, 135
pleasure/pain 36, 91–93, 113–22, 125,
130–31, 134–35
virtue/vice 113, 122–25, 126, 127, 130, 131,
134–35
Audi, Robert 107
Austin, J.L. 6
autonomy, value of 19, 35–36, 139–53
deliberated 142, 144–45, 147–49
balancing principles 41–45, 52–54
Rashdall on 43, 49–50
beauty 95–99, 101, 116, 196–97
admiring contemplation of 4, 95–99, 101
benevolence 74, 76–78, 86, 113, 122, 131, 193
Bentham, Jeremy 67–69, 113, 115, 131
Bishops, US Catholic 242
Blackstone, Sir William 127–28
bonum progressionis 102
Bradley, F.H. 37–38
Brave New World 38, 114, 141
Brentano, Franz 109, 122, 193
Broad, C.D. 3, 5, 8, 13
business, value of 135, 188, 196
callousness 102, 123–25
capital punishment. *See* punishment, capital
castle doctrine 255
Chisholm, Roderick M. 102, 105, 109
Christopher, Paul 261
coercion 20, 141, 150–51, 202–7, 213, 254–55
coherentism 14, 21–22
commonsense morality xiii, 3–8, 13, 15, 17, 20,
169, 178, 180–84, 219, 227, 256, 258
communitarianism 220–24
comparison of values. *See* values, comparison of
compassion 7, 18, 76–79, 81–82, 85–86, 90,
105–7, 113, 116, 122, 124, 131, 198
concentration, costs of 46–52
consequentialism 4, 7, 15, 18–19, 35–36, 37, 54,
150, 215, 242, 254
and population ethics 27, 29–30, 32, 35–36
and punishment 128, 201–2, 205, 212
satisficing 21, 215–18
and virtue 77–78
courage 74, 178–80
culture, survival of 225–29, 255
Czechoslovakia 239, 250–51, 256
Dancy, Jonathan 57, 60–61
Darwin, Charles 49–50
da Vinci, Leonardo 41–42, 44, 52
death penalty. *See* punishment, capital
deliberation 142, 144–45, 147–49, 190
deontology. *See* nonconsequentialism
desert xii–xiii, 6–7, 11–12, 14–15, 21, 101–12,
116, 125–29, 154–77
economic 108, 154, 166–67
holistic 154, 157–60, 161–66, 170–77
individualistic 154, 157–60, 161–66, 170–77
Moore on 5, 11, 103–4
moral 101–12, 125–30, 155–61, 167

- positive vs. negative 131–32, 135
 and proportionality 7, 12, 14, 125–26, 128–29,
 132, 155–60, 161–66, 168, 170–72, 174
 Rawls on 154, 167, 169, 175–76
 retributive 5, 11–12, 103–4, 107, 108, 112,
 127–29, 154, 160–66, 167, 169, 171, 175,
 176, 201–2, 204–5, 211–13
 Ross on 103–4
 de Sousa, Ronald 57, 60–61, 64
 deterrence 128–29, 201–2, 205, 211–13,
 244–45, 247, 253
 dilettante's disadvantage 46–52
 discrimination condition 239–40, 246, 256,
 261
 doing/allowing distinction 20, 35–36m 130,
 259–60
 Dolinko, David 162, 164–65
 double effect, doctrine of. *See* intending/
 foreseeing distinction
 Douglas, William O. 163–64

 envy 82, 89–90, 111, 122, 131, 161
 Epicurus 114
 equality 14–15, 16, 21, 71–72, 116–17, 119, 129,
 160
 evil, problem of 131, 135
 Ewing, A.C. 126–29
 experience machine 133, 188, 190
 expressivism 5, 7–8, 20

 failure 38, 75, 89, 133–35. *See also* achievement,
 value of
 Falklands War 244, 246, 256
 false belief 75, 89, 109, 133–34, 188. *See also*
 knowledge, value of
 Feinberg, Joel 155, 167, 170–73
 Feldman, Fred 103–4
 flourishing xi, xiii, 13, 18–19, 181
 foundational moral theory. *See* structural vs.
 foundational moral theory
 Frankena, William K. 95, 97, 99, 101–2
 freedom. *See* autonomy
 friendship 5, 37–39, 59, 90, 105, 231

 games 41, 44, 135, 171, 185–98
 definition of 9, 186–87
 Gauthier, David 13
 Geneva Conventions 240, 246, 248–49
 Gewirth, Alan 13
 Gulf War (1991) 238, 243–45, 247–49, 256–57,
 261–62

 Hamilton, Sir William 43
 happiness 7, 34, 36, 37–38, 63–64, 69, 101–3,
 108, 113–14, 121–22, 130, 167, 215

 Hare, R.M. 12
 hedonism 38, 67–70, 73, 74, 87, 188
 Hegel, G.W.F. 37
 Hiroshima 239
 human nature 38–39, 221
 Humboldt, Wilhelm von 43–44, 52, 139–40,
 147–50, 152
 Hurley, Susan 57

 inequality. *See* equality
 intending/foreseeing distinction 20, 130,
 239–40, 259–60
 intervening agency 247–51, 263–64
 intrinsic value. *See* value, intrinsic
 intuitions, moral 3–4, 11, 13, 19, 242, 256
 Iraq War (2003) 238, 245, 247–48, 249

 Johnson, James Turner 242
 jokes 90–91
 just cause condition 239, 242–47, 251–52,
 256, 263
 humanitarian just causes 239, 244, 252
 sufficient vs. contributing just causes 243–46
 just war theory 238–64
ius ad bellum 239, 240–41, 242–48, 250–56,
 258–60, 262–63
ius in bello 239–41, 245–51, 256–62

 Kagan, Shelly 60–61, 67
 Kant, Immanuel xi, 13, 19, 25–26, 34, 111, 130,
 195, 212
 Kasparov, Gary 197
 Kafka, Gregory S. 121
 Kekes, John 57, 60–61
 Kierkegaard, Soren 195
 killing 8, 11, 20–21, 35–36
 in self-defense 205–7, 209–10, 214, 255
 in war 239–40, 243, 246, 248, 251–55, 256–62
 Kissinger, Henry 262
 knowledge, value of 4, 7, 37–39, 60, 66, 68–69,
 74–75, 80, 100–02, 109, 133–35, 146–49,
 188–90
 Aristotle on 37–38, 189
 comparative value of 74–75, 78, 81, 83–84,
 133–34
 degrees of value in 38–39, 44, 54, 133–34,
 146–47, 188–90
 and desert 107–8
 moral 4, 9
 pursuit of 18, 44, 76, 83–84, 134, 193
 of self 44, 146
 value of as conditioned on pleasure 95,
 97–98, 101
 Korsgaard, Christine M. 13, 18, 98, 101–3
 Kosovo War 238, 252, 253–54, 261–62

- Lackey, Douglas P. 242
 Lamont, Julian 173
 last resort condition 239, 240–41, 243
 Leibniz, G. W. 25–26, 34, 37
 Lemos, Noah M. 96, 105, 107–8
 Leslie, John 36
 liberalism 139–41, 149–53, 252
 Locke, John 201–2, 206, 211
 love 4, 114, 135, 231–32, 235
 luck. *See* responsibility
- MacIntyre, Alasdair 223–24
 malice 7, 9, 74, 76–77, 82–83, 86–93, 105–12,
 122–25, 131, 178–80
 Mandela, Nelson 197
 Marcus Aurelius 52
 Marx, Karl 37, 43, 145–46, 195, 197–98, 221
 Mayerfeld, Jamie 113–15, 118–20, 125, 131
 McKenna, Joseph C. 243, 251
 McKim, Robert 243
 McMahan, Jeff 219, 225–26, 233, 243, 254
 McTaggart, J.M.E. 3, 4, 27
 mere-addition objection 33–35
 Mill, John Stuart 38, 95, 139–42, 144–50, 152
 Miller, David 173, 220
 minimum necessary condition. *See* necessity
 condition
 monism about values 55, 60–73
 Moore, G.E. xi, 3–4, 13, 37, 73, 98, 99–100
 on desert 5, 11
 on intrinsic value 65, 96
 on organic unities 5, 8, 11, 94–97, 101, 103,
 104
 on pleasure and pain 113–15, 118–20, 131
 on pluralism about values 73, 135
 on virtue 18, 81–82, 85, 109, 122, 193, 198
 Moore, Henry 197
- Nagel, Thomas 19
 narcissism. *See* self-indulgence, moral
 nationalism 219–37, 258–61
 and embedded selves 220–24
 full-blooded 228
 and history 229–37
 and impersonal goods 225–29
 universalist 219–20, 222–24
 necessity condition 206–7, 209–14, 240–41,
 246, 257
 Newman, John Henry Cardinal 78
 Nietzsche, Friedrich 14–15, 22, 37, 43, 195–96,
 221
 nonconsequentialism 4, 6, 11, 13, 15, 20–21, 78,
 127–28, 130, 201
 nonnaturalism 3, 20
 Norman, Richard 252–56
- Nozick, Robert 12–13, 17, 54, 103, 109, 133,
 168, 170, 172, 188, 202–3, 206
 Nussbaum, Martha C. 60, 64–65
- O'Brien, William V. 249
 optimism vs. pessimism 131, 135
 organic unities, principle of 5, 8, 20, 94–112
 and achievement 95, 97–98, 100–1
 and aesthetic contemplation 95, 97–101
 conditionality interpretation of 97–112
 and desert 11, 101–12
 generic 94–95, 99, 102
 holistic interpretation of 96–104, 112
 and knowledge 95, 97–98, 100–1
 Moore on 5, 8, 11, 94–97, 101, 103, 104
 and virtue and vice 101–2, 104–12
 ordinary-language philosophy 6
 Osirak raid 244
- pain, disvalue of 5, 19, 36, 58, 75, 103–7, 109,
 111, 198, 247
 comparative disvalue of 58, 78–83, 85–93,
 113–22, 125, 131, 133–34
 and desert 94, 103–4, 106–7, 126, 155, 167
 increasing marginal disvalue of 92
 linear view of 91
 and population ethics 36
 and virtue and vice 58–59, 76, 90, 94, 104–7,
 111
- Parfit, Derek xi, 27, 95, 119, 121, 160, 197
 partiality 5, 8, 219–37, 258–60
 and history 229–37
 toward culture 235–36
 toward intimates 219–20, 226–28, 230–35,
 258–60
 toward nation and conditionals 219–37,
 258–60, 261–62
 objective vs. subjective basis of 230–31, 235
 racial 219, 228–29, 230, 233–34
- particularism 9, 221, 223–24
 patriotism. *See* nationalism
 perfection
 physical 38, 40, 43–44
 practical 38–39, 40
 theoretical 38–39, 40
 perfectionism 10, 12, 15, 37–38, 41, 74–75,
 139–40, 150, 152, 185
 Aristotelian 38–39
 and balancing principles 41–45, 49–51,
 52–54
 cultural 221–23
 and liberalism 149–53
 and mere additions 34–35, 37–38
 permissions, agent-relative 13, 15, 20, 218
 persons, separateness of 12–13, 14–15, 53–54

- Plato 18, 37, 51, 114, 129
- pleasure, value of 4, 9, 38, 41, 58–73, 74–75, 109, 114, 129, 133, 188, 194, 215, 232, 246
- Benthamite vs. non-Benthamite views of 67–72
- comparative value of 36, 75–76, 78–83, 86–93, 113–22, 125, 130–31, 133–34
- as a condition of value 95, 97–98, 101
- and desert 101–4, 105–12, 155, 167
- diminishing marginal value of 91–93, 118–19
- and population ethics 36
- and virtue and vice 57–58, 76, 178–79, 193
- pluralism about values 10, 11, 14, 39, 55, 60–73, 150–52
- population ethics xii, 22–36, 121–22
- average principle 27–31, 34, 35
- critical-level view 121–22, 124
- total principle 27–31
- variable value view 26–27, 29–36
- priggishness. *See* self-indulgence, moral
- priority view 21, 72, 119–20, 160
- proportionality xii–xiii,
- and desert 7, 12, 14, 125–26, 128–29, 132, 155–60, 161–66, 168, 170–72, 174
- and punishment 127–29, 161–66, 205–7, 209–14
- and regret 56–69
- and self-defense 205–7, 209–14, 241–42, 248, 254–55
- and virtue xii, 56–61, 68–69, 82–84, 86–87, 101–2, 122–23, 131
- and war 238–64
- punishment xii–xiii, 5, 11–12, 16, 202–14, 244, 246–47
- capital 162–66, 171, 201, 208, 212–13
- and retributive desert 5, 103–4, 108, 112, 125–29, 132, 154, 160–66, 167, 169, 171, 174–76, 201–2, 204–5, 211–13
- and rights 201–14
- Pythagoreans 129
- Quebec 225–27, 235–36
- Quine, W.V.O. 6
- Quinn, Warren 213
- rape 20, 162, 164, 214, 252, 255
- Ramsey, Paul 249
- Rashdall, Hastings 3, 4, 19, 37, 73, 191–93
- on balancing principles 43, 49–50
- on virtue 79–81, 122
- Rawls, John 6, 11, 13–14, 15–16, 19–20, 38
- on desert 154, 167, 169, 175–76
- on utilitarianism 12, 15, 53–54
- realism, moral. *See* nonnaturalism
- reasonable hope of success condition 239–40
- reflective equilibrium. *See* coherentism
- Regan, Richard J. 243
- regret, rational 55–73
- concentration view of 56–57
- modal conditions on 59–60
- and monism vs. pluralism about values 55, 60–73
- proportionality view of 56–61, 68–69
- relativism, moral 20
- repugnant conclusion 27, 32, 35, 121–22
- research. *See* knowledge, pursuit of
- respect for persons 12, 15–16, 111, 228, 237
- responsibility 168–69
- revenge 103
- retributivism. *See* desert, retributive
- rewards xii, 15, 103, 108, 125–26, 132, 156, 161, 167–71, 245
- rights 15, 21, 220, 228, 237, 239, 241, 244, 251–52, 258
- to collective political self-determination 252–56
- comparison of 203–4, 207–9
- to liberty 152, 203–5, 208–10
- natural 6–7, 201–4
- to punish 201–5, 210–14
- of self-defense 202, 204–7, 241–2
- Rodin, David 252–56
- Rose, Pete 192–93
- Ross, W.D. xi, 3, 73, 84
- on desert 103–4
- on virtue 4, 18, 78, 88, 109, 122, 179–82, 193
- Rwanda 252
- Saddam Hussein 249, 262
- sadism. *See* malice
- Sadurski, Wojciech 173
- sanctions, economic 249, 251
- satisficing 21, 215–18
- self-defense 202–3, 205, 206, 209, 213–14, 241–42, 248, 252–56
- self-indulgence, moral 18–19, 83–84
- self-interest 4, 12, 13, 18
- selfishness 101–2, 131
- separateness of persons. *See* persons, separateness of
- Scheffler, Samuel 13, 15, 154, 160, 175–77
- Schiller, Friedrich 195
- Schopenhauer, Arthur 231
- self-referential altruism. *See* partiality
- Shakespeare, William 231
- Sher, George 11, 173
- shields, voluntary and involuntary 248–50
- Sidgwick, Henry 3, 4, 5, 19, 113, 115, 131
- Simmel, Georg 195
- situationist social psychology 183–84

- Slote, Michael 107, 215–18
 sloth 123
 species, animal 25–26
 sports. *See* games
 Stocker, Michael 57, 60–63, 67
 structural vs. foundational moral theory xi–xii, 4–5, 10, 11–22
 suicide bombers 248, 250, 258
 Suits, Bernard 9, 185–87, 191–92, 195, 197–98

 Tamir, Yael 232
 Taylor, Charles 225
 Thomson, Judith Jarvis 259–60
 threats. *See* coercion
 Truman, Harry S. 239

 upper limit qualification. *See* proportionality
 utilitarianism 4, 6, 7, 34–35, 37, 73, 113, 201, 258
 critical-level 121–22, 124
 negative 114
 Rawls on 12, 15, 53–54

 value, intrinsic 4, 6–7, 105, 167
 definition of 64–66, 96–99
 vs. instrumental value 64, 140, 197–98
 Moore on 65, 96, 103–4
 moral 80–81
 values, comparison of 4, 10, 11, 39–54, 74–93, 104–5, 112, 113–35, 151, 164–65, 251–63
 balancing 41–54
 constant 40–41
 lexicographic 39–40, 78–79, 88, 151
 pure specializing 50
 van den Haag, Ernest 164
 vice, disvalue of 5, 8, 74, 105–6, 107–8, 111, 113, 155, 161, 167

 comparative disvalue of 75, 78–93, 116, 122–25, 130, 131
 dispositional vs. occurrent-state view of 178–84
 and proportionality 101–2
 recursive account of 75–78, 109–10, 122–23, 179, 193
 See also virtue, value of
 Vietnam War 249, 251, 261
 virtue, value of xii, 4, 6, 8–9, 13, 18, 57, 74–93, 101, 107–8, 109–10, 113, 122–25, 128–29, 134–35, 155, 167, 181
 Aristotle on 55, 178, 181, 183, 193
 comparative value of 75, 78–93, 116, 122–25, 130, 131
 dispositional vs. occurrent-state view of 178–84
 Moore on 18, 81–82, 85, 109, 122, 193, 198
 positive vs. negative forms of 131, 135
 and proportionality xii, 56–61, 68–69, 82–84, 86–87, 101–2, 122–23, 131
 Rashdall on 79–81, 122
 recursive account of 75–78, 109–10, 122–23, 179, 193
 Ross on 4, 18, 78, 88, 109, 122, 179–82, 193
 virtue ethics 13, 18–19, 122, 179–81, 183

 Walzer, Michael 240–41
 war, ethics and 227, 232, 238–64
 weakness of will 60–61, 143
 welfarism 74–75
 well-roundedness 41–54
 Wiggins, David 60
 Williams, Bernard 57, 83–84
 Wittgenstein, Ludwig 6, 8–9, 185
 Woods, Tiger 197
 Wordsworth, William 49–50